A Hypercolumn Based Stereo Vision Model

By

Lam Shu Sun

A thesis submitted to the

Department of Computer Science

The Chinese University of Hong Kong

in partial fulfillment of the requirements

for the degree of

Master of Philosophy

May 1993

Acknowledgments

I would like to express my gratefulness to the many people who have contributed to this work. Dr. K. H. Wong who gave guidance, and in depth discussions on this project. S. H. Or who helped a lot in interpolation of the figures. I wish to thank C. M. Hui for his help with preparing this thesis.

Finally I would like to thank the colleagues in the Department of Computer Science, The Chinese University of Hong Kong for their assistance in this work and providing the atmosphere necessary for the research.

Abstract

Stereo vision is a method to extract the depth information from two different viewpoints of the same scene. The critical problem is how to locate the corresponding points from the left and right hand images. Many algorithms have been proposed to solve this problem, however, their results were poor in handling natural scenes.

Recent progress in the study of human vision has renewed the interest in computer vision. A new stereo vision model based on the psychophysical evidences in human visual system is proposed in this thesis. The psychophysical based stereo vision model (PSVM) uses a special data structure called hypercolumns to handle the image features. It also uses the fusional method similar to that in human to solve the stereo correspondence problem. And this binocular processing can be implemented in parallel.

Our approach is implemented and tested on a number of synthetic images as well as natural images. The results are satisfactory.

Contents

Chapter 1

Introduction: Binocular Depth Visual Perception of Human **1**

1.1 Introduction 1

1.2 The visual pathway 2

1.3 The retina 3

1.4 The ganglion cells 5

1.5 The lateral geniculate nucleus 7

1.6 The visual cortex 8

1.6.1 The cortical cells 8

1.6.2 The organization of the visual cortex 9

1.7 Stereopsis 11

1.7.1 Corresponding retinal points 12

1.7.2 Binocular fusion 14

1.7.3 The binocular depth cells 14

1.8 Conclusion of chapter 1 15

Chapter 2

Computational Stereo Vision **16**

2.1 Stereo image geometry 16

2.1.1 The crossed-looking geometry 17

2.1.2 The parallel optical axes geometry 19

2.2 The false targets problem 20

2.3 Feature selection 21

2.3.1 Zero-crossing method 21

2.3.2 A network model for ganglion cell 24
2.4 The constraints of matching 28
2.5 Correspondence techniques 29
2.6 Conclusion of chapter 2 29

Chapter 3

A Hypercolumn Based Stereo Vision Model 30
3.1 A visual model for stereo vision 30
3.2 The model of PSVM (A Computerized Visual Model) 32
3.3 Local orientated line extraction (Stage 1 of PSVM) 34
3.3.1 Orientated line detection network 35
3.3.2 On-type orientated lines and off-type orientated lines 37
3.4 Local line matching (Stage 2 of PSVM) 38
3.4.1 Structure of hypercolumn in PSVM 39
3.4.2 Line length discrimination model (Part of stage 2 of PSVM) 41
3.4.3 Orientation-length detector 42
3.4.4 Line length selection 45
3.4.5 The matching model 46
3.4.6 Fusional area in PSVM 48
3.4.7 Matching mechanism 49
3.4.8 Disparity detection 50
3.5 Disparity integrations (Stage 3 of PSVM) 53
3.5.1 The voter network 54
3.5.2 The redistributor network 55
3.6 Conculsion of chpater 3 57

Chapter 4

Implementation and Analysis **58**

4.1 The imaging geometry of PSVM 58

4.2 Input 59

4.3 The hypercolumn construction 59

4.4 Analysis of matching mechanism in PSVM 59

4.4.1 Fusional condition 61

4.4.2 Disparity detection 61

4.5. Matching rules in PSVM 63

4.5.1 The ordering constraint 63

4.5.2 The uniqueness constraint 64

4.5.3 The figural continuity constraint 64

4.5.4 The smoothness assumption 65

4.6. Use multi-lengths of oriented line to solve the occlusion problem 66

4.7 Performance of PSVM 67

4.7.1 Artificial scene 67

4.7.2 Natural images 71

4.8 Discussion 83

4.9 Overall conclusion 83

Appendix: Illustration example **85**

References **91**

Chapter 1

Introduction: Binocular Depth Visual Perception of Human

1.1 Introduction

Computer stereo vision is a methods to extract depth information of the world. The depth can be recovered from two images of a scene which are taken from two slightly different views by measuring the differences, also called disparities, of the images. The problems of computer stereo vision can be separated into three parts:

1) Feature selection which is the problem of selecting reliable features for matching.

2) Correspondence problem which the matching of features to extract the disparities of the two images.

3) Interpretation problem which is the use of disparity information to recover the depth of the scene.

The correspondence problem is a critical problem in stereo vision because it is not a one-to-one mapping between two images. To obtain an accurate depth map, the false target problem must be solved first (see section 2.2).

Obviously, there is considerable commercial and engineering interest in the development of efficient stereo-algorithms. In general, stereo vision models can be divided into two types:

1) Computer based stereo vision models which consider stereo vision problem as an information computational problem [1, 2, 11, 15, 20, 22, 28, 29].

2) Psychophysical based stereo vision models which use the psychophysical findings about human binocular vision to solve the stereo vision problem [10, 12, 13, 22, 25].

By using the evidences in visual system of human, an implementation of new stereo vision called PSVM (Psychophysical based Stereo Vision Model) is presented. It uses a data structure called hypercolumn structure (see section 1.6 and section 3.4.1) to store selected features and results to achieve the binocular processing. PSVM also uses a special feature detector called simple cell (see section 1.6 and section 3.3) to extract features for stereo correspondence. Furthermore, it uses the fusional method similar to that in human to solve the stereo correspondence problem.

In PSVM, feature extraction can be implemented in parallel for feature detectors are independence. Moreover, the feature matchings in stereo correspondence are also in parallel. It is possible to implement a faster stereo vision system by using PSVM structure. With the fusional method, PSVM allows misalignment which is important for a stereo vision system for it is very difficult to obtain a pair of stereo image with perfect alignment.

In this thesis, a brief description about binocular depth visual perception of human and computational stereo vision will be introduced in Chapter 1 and Chapter 2. The model of PSVM will be discussed in Chapter 3. Finally, the implementation and analysis, overall conclusion, are given in the Chapter 4.

1.2 The visual pathway

The human visual system is a complex one. Before we can perceive a scene, the light reflected from it must travel a long and complex path through the eyes to the visual cortex, in which, information of the scene, for example, depth, color and forms, are obtained.

When a human is looking at an object, his/her eyes are focused directly towards the same point. Light goes through the cornea (Figure 1.1), and some passes through the pupil and crystalline lens until it strikes the retina, where it forms an optical image and that transmits along the optic nerve [9]. These electrical impulses arrive at the lateral geniculate nucleus, where the impulses are retransmitted until they reach the visual cortex at the back of the brain. Here, the real world image is decomposed and

the information is ready for other neurons in the cortex. This visual pathway is shown in Figure 1.2. They work together to give us the sensation of a single external world.

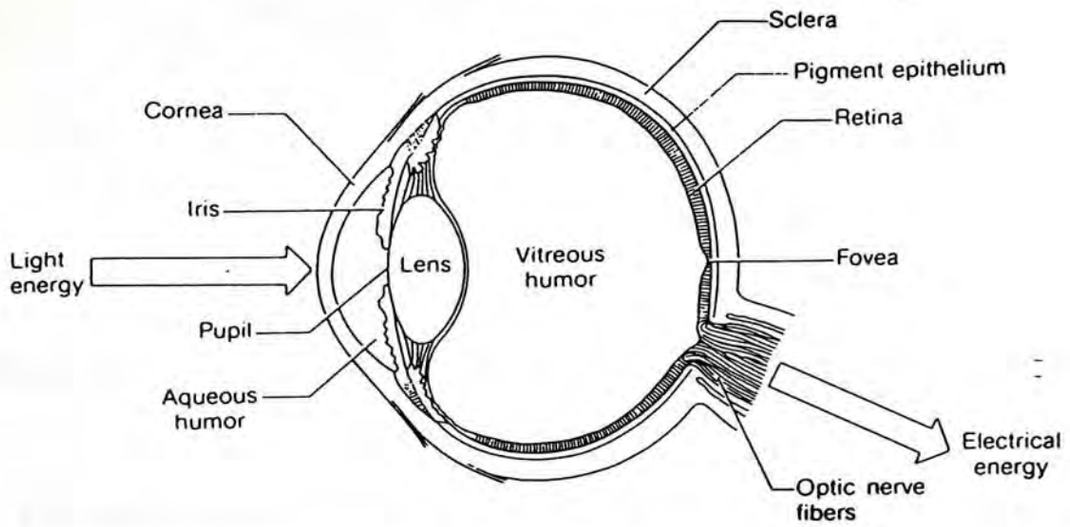


Figure 1.1 A cross section of the human eye. (From Goldstein, 1989.)

1.3 The retina

The main structure of the retina is formed on a vertical basis [14] (Figure 1.3). The retinal receptors are faced away from the light so that it must pass through the ganglion, bipolar, and horizontal cells before reaching the receptors [9, 14]. There are two different types of photoreceptors in the retina: rods and cones. The rods are longer and are more effective in dim light, whereas the cones are shorter and are more effective in bright light [9, 14]. These rods and cones are not evenly distributed in the retina. The fovea contains only cones, while the rods are rich in the peripheral retina. The fovea is a small area that is located directly on the light of sight, so at any time when we look directly at an object, the image falls on the fovea. However, the fovea is so small that only objects with a visual angle of about 0.5 degree can fall completely on the fovea [9]. As the receptors are packed closer together in this area, the image is more precisely processed.

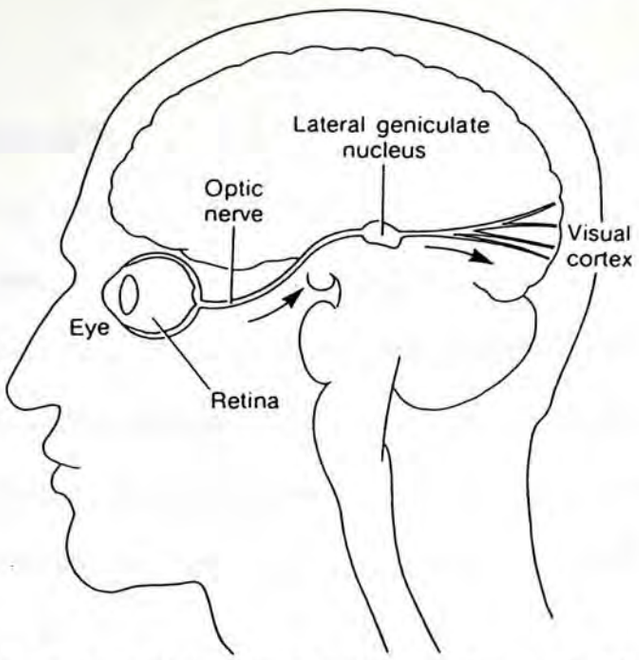


Figure 1.2 A side view of the visual pathway. (From Goldstein, B. E., 1989.)

The rods are rich in the peripheral retina. There are 120 million rods and about 6 million cones in the retina. The ratio is 20 to 1 [9].

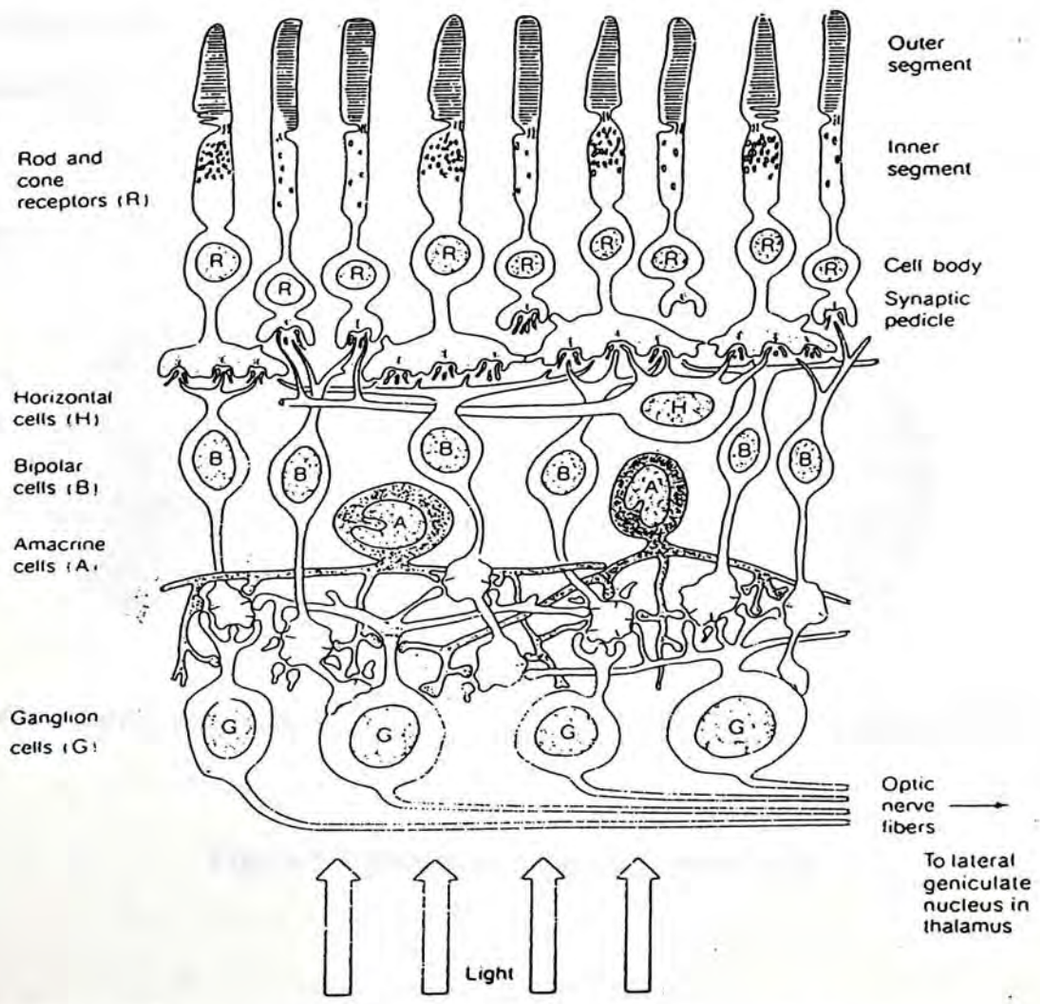
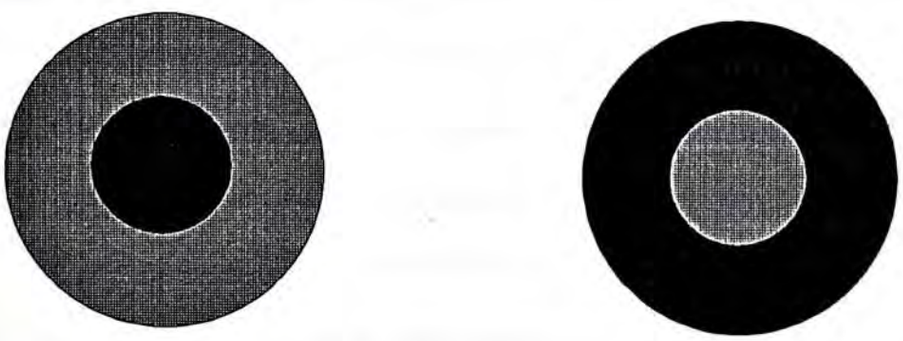


Figure 1.3 Cross section of the primate retina. (From Goldstein, 1989.)

1.4 The ganglion cells

In Figure 1.3, a retinal ganglion cell may be excited or inhibited by a receptor. The area on the retina, covered by the receptors which supply particular ganglion cell, is called the receptive field of that ganglion cell. There are two types of the receptive fields found in the retinal ganglion cells. One is the on-center receptive fields which have excitatory centers and inhibitory surrounds. The other is the off-center receptive fields which have inhibitory centers and excitatory surrounds. These two types of receptive fields are shown in Figure 1.4. The on-center receptive fields give the best response when a spot of light falls onto the center of that receptive field. The off-center receptive fields act exactly in the opposite manner [9, 14]. A neural circuit that can result in an on-center receptive is shown in Figure 1.5. In this circuit, excitatory synapses are represented by ">", and inhibitory synapses are represented by "T" [9]. The receptive field of ganglion cells can be described as a DOG (Difference Of Gaussians) [5, 19, 23]. Its size determines the spatial resolution. If an on-center receptive field and an off-center receptive field work together, they can be served as an edge operator [23].



On-center receptive field Off-center receptive field

Figure 1.4 Receptive fields of ganglion cells

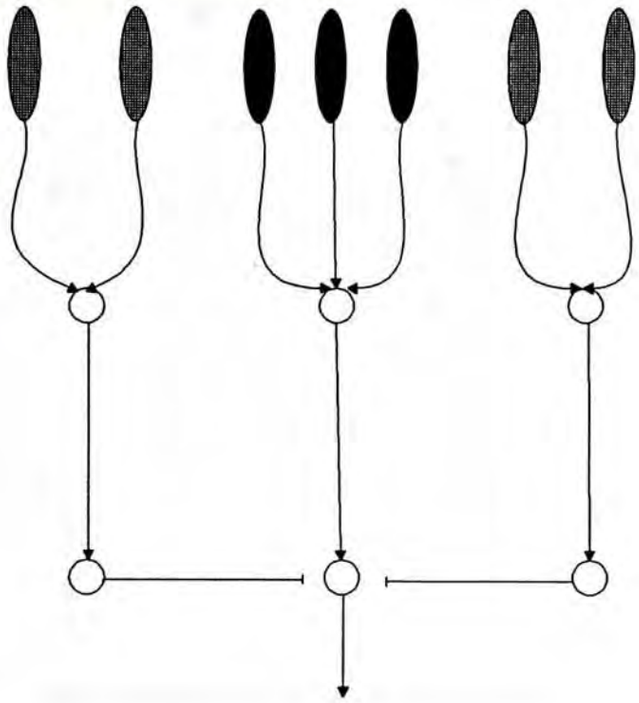


Figure 1.5 A neural circuit that would result in a on-center receptive field. (From Goldstein, 1989)

1.5 The lateral geniculate nucleus

The ganglion cells terminate at the lateral geniculate nucleus (LGN). The receptive fields of the LGN are similar to those of the ganglion cells. Many neurons in the LGN play an important role in color vision [9]. However, their properties in color vision are not concerned in this research and will not be discussed. The retina is a two-dimensional surface and the electrical impulses travel in the optic nerve to the LGN, which, however, is a three-dimensional structure [9, 14]. The LGN is organized into six layers. The information from the retina is organized within these six layers (Figure 1.6). Each layer receives input from only one eye. Layers 2, 3 and 5 receiving input from the ipsilateral eye and layers 1, 4 and 6 receive input from the contralateral eye. The map of the retina in the LGN is a topographic map. The location relative to one another of the receptive fields on the retina are found next to each other in the LGN (Figure 1.6). Thus, any small region of the retina is represented in three layers, alternating with similar projections from the other eye [9, 18].

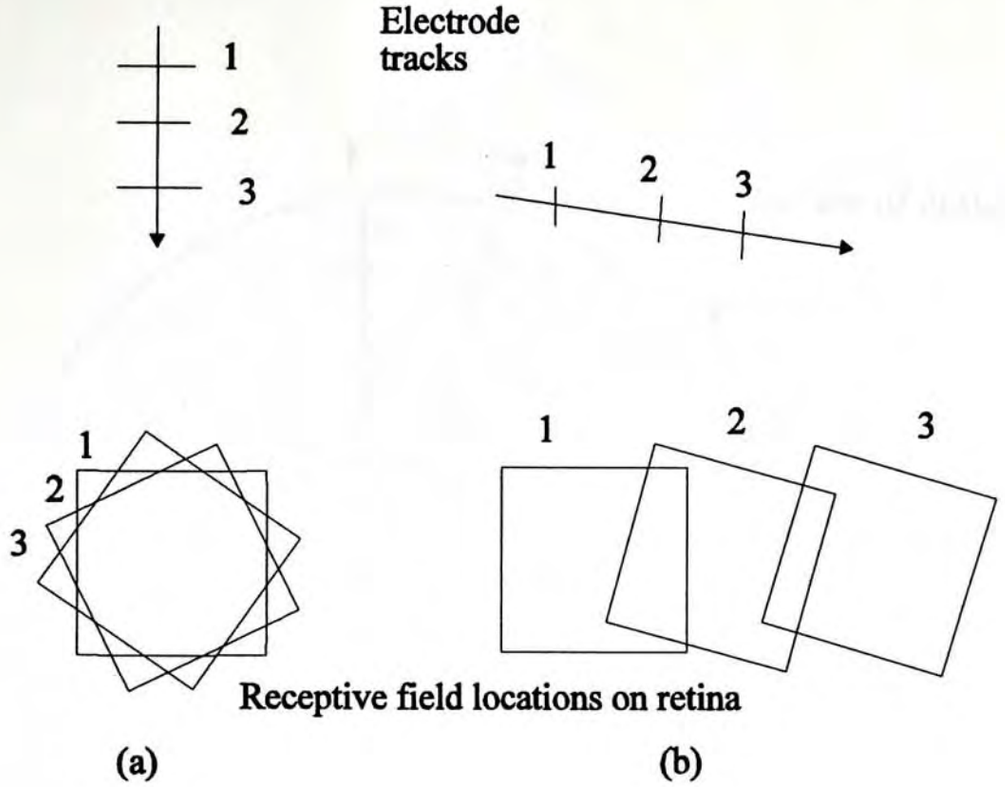


Figure 1.8 Loation columns in the visual cortex. (From Goldstein, 1989)

The oriented lines in the visual cortex are also organized into columns, the orientation columns. Neurons in the same orientation columns have the same preferred orientation. The adjacent orientation columns have similar, but with slightly different orientations [9] (Figure 1.9).

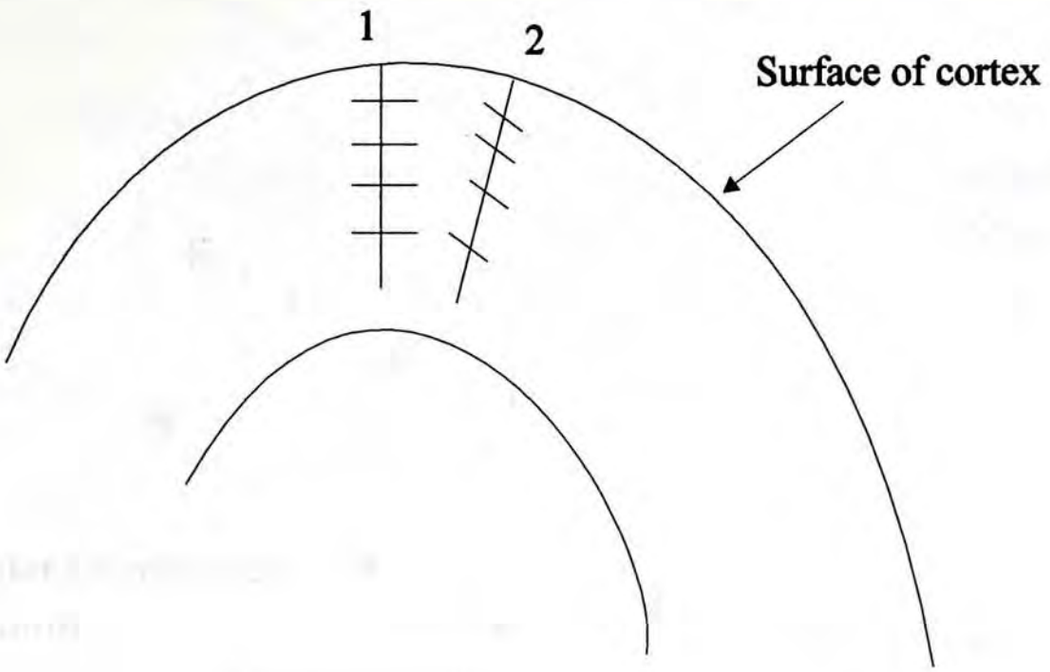


Figure 1.9 Oreintation columns in the visual cortex.

The neurons are not only organized into the location columns and orientation columns, but also are organized into the ocular dominance (left or right view) columns. In the same ocular dominance columns, neurons have the same preferential response to one eye [9, 14, 18].

According to Hubel and Wiesel, these three types of columns can serve as a processing module for a specific area of the retina [18]. They called this module as a hypercolumn. This is shown in Figure 1.10. The shape of the hypercolumn is like a cube. The location column, the orientation column and the ocular dominance column intersect at right angles. This model can represent any line or curve in the visual scene. A long continuous line in retina will be described as discrete short line segments in this module [18].

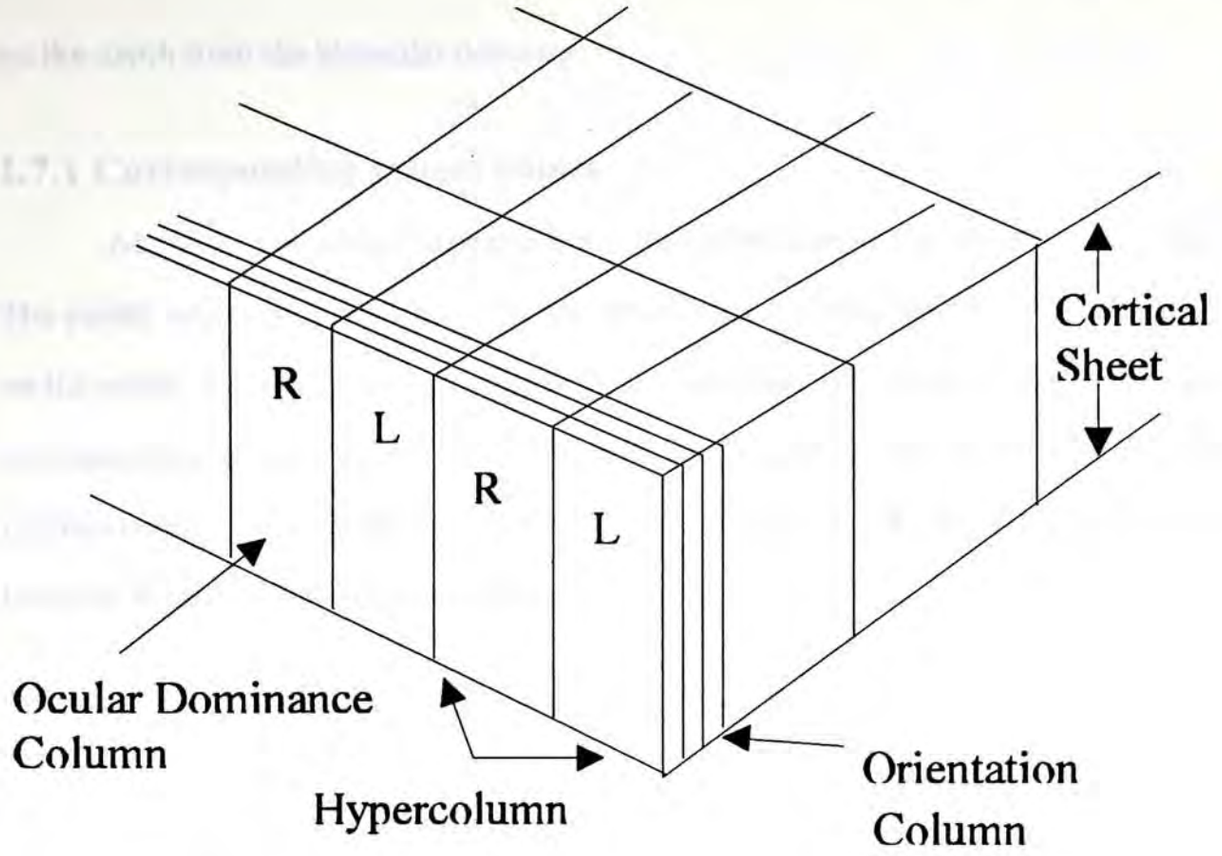


Figure 1.10 A diagrammatic picture of hypercolumns.

In addition to the above columns, another kinds of column structures, disparity columns [6], color columns and contrast (on/off) columns [21], are also found in the visual area. It seems that column structure is the main structure of the visual cortex.

1.7 Stereopsis

We can perceive the depth in a scene although it is projected onto the two-dimensional surface of our retinas. The depth can be obtained by many cues, for examples, oculomotor cues, pictorial cues, motion-produced cues and binocular disparity. The oculomotor cues are the sense of the position of our eyes and the tension in our eye muscles. The pictorial cues are obtained from a still picture, such as comparing the size of the objects or judging the overlapping of the objects. The motion-produced cues are depended on the movement of the observer or the objects' movement in the environment. The binocular disparity is obtained from two slightly different images of a scene formed on each retina [gol89]. We will focus our attention

on the depth from the binocular disparity.

1.7.1 Corresponding retinal points

As mentioned above, disparities are the differences of the two retinal images. The points which have zero disparity are called corresponding points. For every point on the retina, there is a corresponding point on the other. The two foveas, F and F', are corresponding points, as shown in Figure 1.11. A and A' and B and B' are also corresponding points. If B and A' are the same image of a scene, the displacement between B and A' is called the disparity of these images.

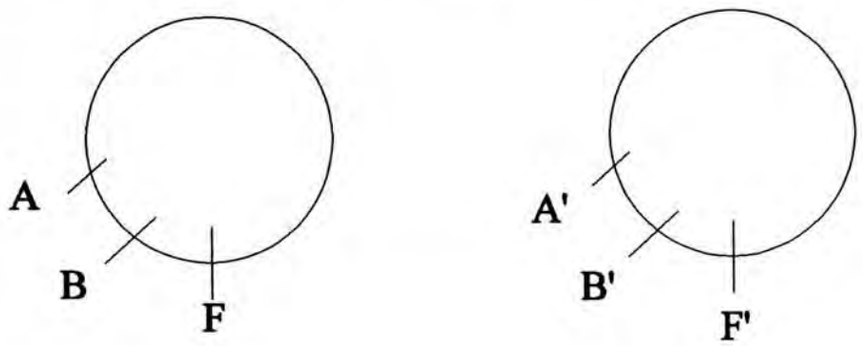


Figure 1.11 Corresponding points on the two retinas.

Given a fixation point, there must be a set of points which have zero disparity. The locus of the zero disparity points is called horoptor. As shown in Figure 1.12, the dashed line is the horoptor. The points, F, A and B, are on the horoptor so that they are the corresponding points and are on the convergence distance. However, the points, C, D and E, are the noncorresponding points. The disparities of C and D are called crossed disparities because these images cross before the convergence distance, whereas the disparity of E is called uncrossed disparity for it would not cross before the convergence distance [34]. Note that if the object is farther from the horoptor, it will have a greater disparity. In Figure 1.12, the disparity of C is the distance between A and B', whereas the disparity of D is the displacement between A and F'. As the disparity of C is greater than D, C is located farther from the horoptor and closer to

viewer. It should be pointed out that the disparity is the relative depth reference to a fixation point. However, it is easy to obtain the distance of an object with a simple calculation from trigonometric law. This will be discussed on Chapter 2.

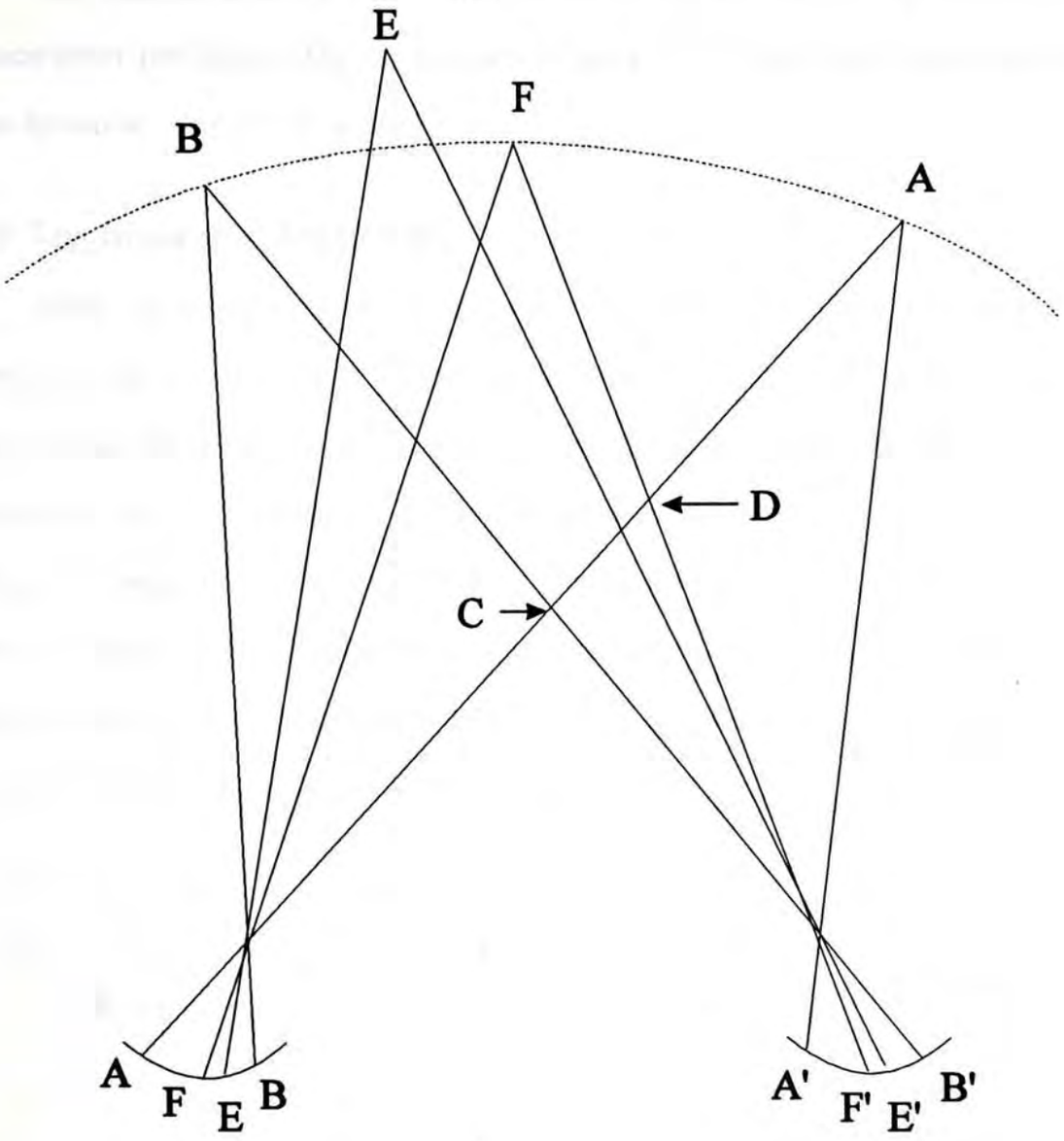


Figure 1.12 The horoptor and the corresponding points.

1.7.2 Binocular fusion

If the images in the two eyes are very similar and their disparity is within a specific range, these images will be fused into a single image (Figure 1.13). This range is known as Panum's area [14, 34, 35]. Within the Panum's area, the similar images are

not only fused into a single image but also remian available in the visual system for the depth perception [34]. Beyond this area, the images are perceived separately by each eye. This perception is called diplopia or double vision [14, 34].

The Panum's area is not of a fixed size. It increases roughly proportional to the distance from the fovea [35], as shown in Figure 1.14. The fusing increases allow fusion for up to about 2^{o} of arc [35].

1.7.3 The binocular depth cells

Many experiments showed that there do exist cells which can detect the disparity in the visual cortex. Barlow, Blakemore and Pettigrew found cells in the cortex of cats that have the best response to the stimuli separated by a specific disparity on the two retinas [3]. Hubel and Wiesel recorded these cells in the visual cortex of monkey [17]. Disparity sensitive neurons have also been found in the visual cortex of sheeps [34]. The existence of binocular depth cells mean that they will be excited by objects at different relative depths if the convergence is fixed (that is, look at a particular point in space and don't move) [9, 30].

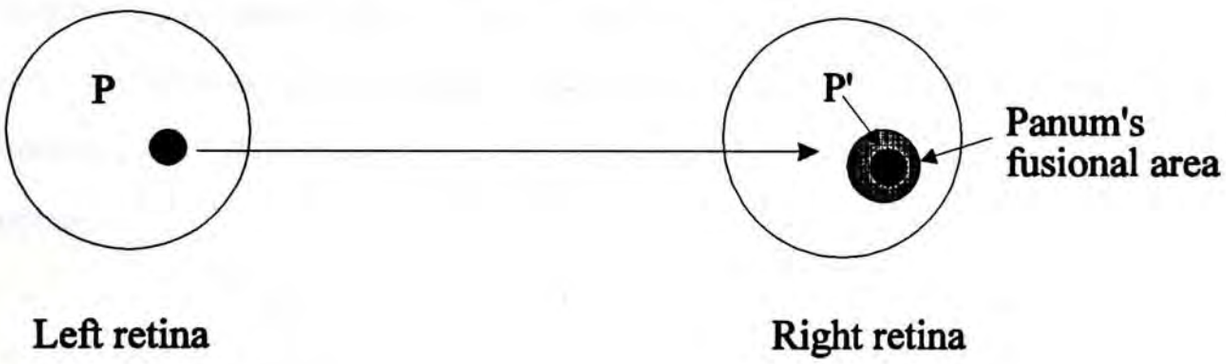


Figure 1.13 Images within the Panum's fusional area will be fused as a single image.

Based on their properties, the binocular depth cells can be divided into three pools: the zero disparity, the crossed disparity and the uncrossed disparity [30]. The neurons will have zero disparity if their receptive fields are in exact correspondence in the two retinas. If their receptive fields have different relative positions, some will have

crossed disparity and the others will have uncrossed disparity. The binocular depth cells having crossed disparity are also called near neurons because they give excitatory responses to the objects nearer than the fixation point, whereas the other ones have uncrossed disparity are called far neurons for their excitation for farther objects than the fixation point [30]. These two kinds of neurons are complement.

It should be pointed out that the disparity sensitive neurons are not evenly distributed. Hubel and Wiesel suggested that the binocular depth cells that have nonzero disparity occur outside the primary visual cortex [17]. There are a lot of neurons in area 17 (primary visual cortex) that are sensitive to zero disparity while compared to the nonzero disparity sensitive neurons. However, in area 18, there are dominated by the nonzero disparity sensitive neurons [8, 17].

1.8 Conclusion of chapter 1

Depth perception includes, matching corresponding points of the images in the two retinas., measuring their disparity, and recovering the 3D structure of the objects [pog84]. The correspondence problem is a main problem in stereo vision. How human to solve this problem is still not known. However, it is clear that a single depth cue is not enough for depth perception. The more cues we have, the better our chances to deduce the three-dimension world from the two dimension images projected on our retinas.

Chapter 2

Computational Stereo Vision

Stereo vision is one of the most important methods for computers to extract depth information from the three-dimensional world. The depth of images is easily measured by our brain. Surprisingly, the development of automatic stereo vision system is still not too successful.

Stereo vision can be broken into three parts: feature selection, stereo correspondence, and disparity interpretation [30, 31]. The stereo correspondence is to match those elements in the two images. The disparity interpretation is the process which maps the disparities into the three-dimensional scene [31]. The stereo correspondence problem has dominated the field of computational and psychophysical researches in stereo vision processing.

2.1 Stereo image geometry

There are two classes of stereo images geometry. The first one is the crossed-looking geometry. The second one is the parallel optical axes geometry. It is assumed that the two cameras are identical and the optical axes of these two cameras are lain on the same plane. Any point in the scene will project to these cameras. The connection of these two image points will produce a line which parallels to the baseline (the distance between the centers of the two lens). Thus, corresponding edges in the two images must lie along the same line, as shown in Figure 2.1. This line is called the epipolar line [2]. The difference between these models is that the two optical axes of the crossed-looking model are not parallel (Figure 2.2), while the other one does (Figure 2.3). The advantage of the parallel optical axes model is that the epipolar will be horizontal.

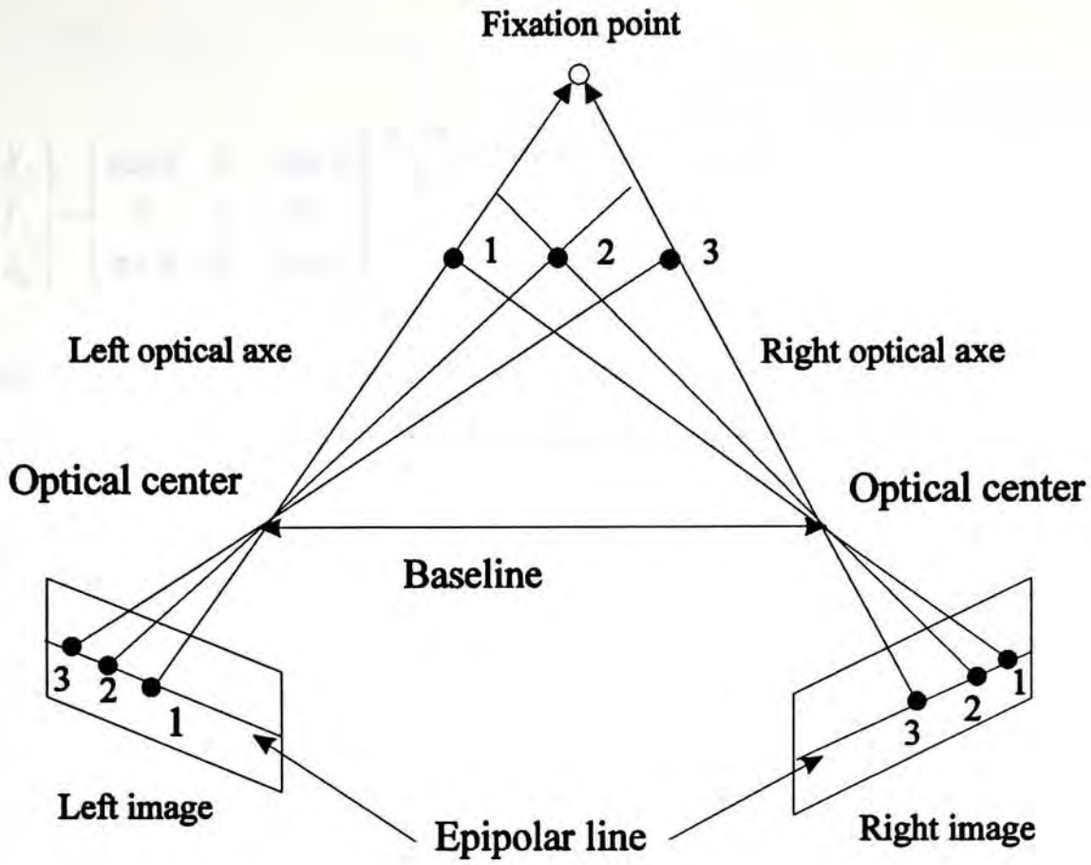


Figure 2.1 Geometry of stereo vision.

2.1.1 The crossed-looking geometry

Assuming that two cameras with the same focal length f and baseline between the centers of the lenses O_1 and O_2 is b (Figure 2.2). The angle between these two optical axes, Z_1 and Z_2, is 2θ. Note that there are two coordinate systems associated with the two cameras, $X_1Y_1Z_1$ and $X_2Y_2Z_2$. Let's define the third coordinate system XYZ, such that Z axis bisects the angle between Z_1 and Z_2. The relations between these coordinate systems are given by

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} \frac{x+b}{2} - f\sin\theta \\ Y \\ Z \end{bmatrix} \tag{2.1}$$

$$\begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} \frac{x-b}{2} + f\sin\theta \\ Y \\ Z \end{bmatrix} \tag{2.2}$$

If point P at (X, Y, Z) projects on the left image at (x_1, y_1) and on the right image at (x_2, y_2), as shown in Figure 2.2, their relationship is given by

$$x_1 = \frac{f X_1}{f - Z_1} \tag{2.3}$$

$$x_2 = \frac{f X_2}{f - Z_2} \tag{2.4}$$

$$y_1 = \frac{f Y_1}{f - Z_1} \tag{2.5}$$

$$y_2 = \frac{f Y_2}{f - Z_2} \tag{2.6}$$

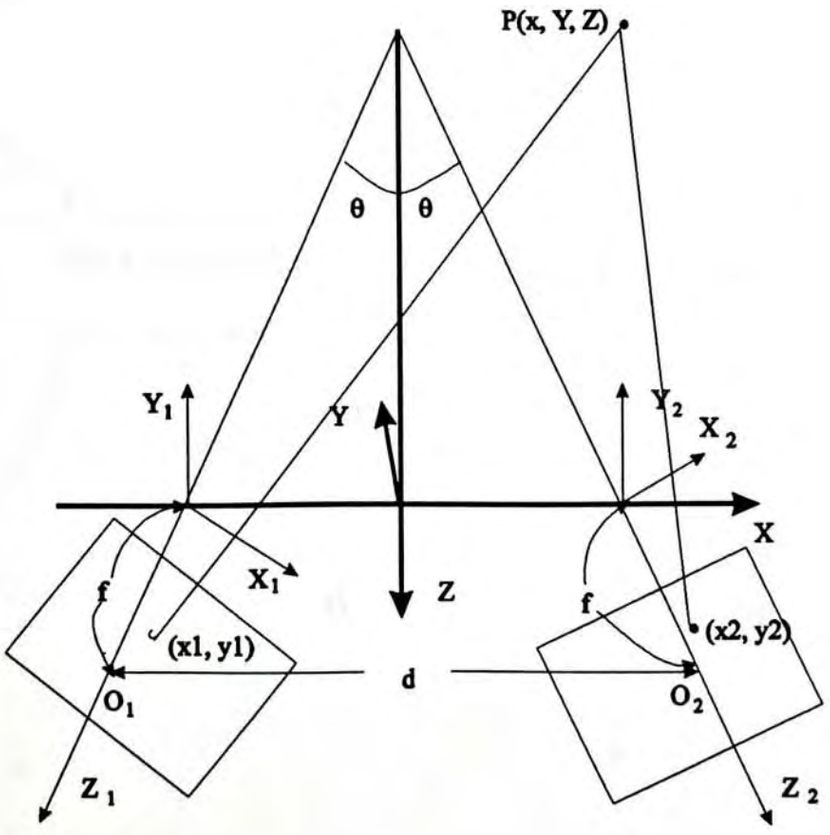


Figure 2.2 The crossed-looking geometry of stereo vision.

2.1.2 The parallel optical axes geometry

In the crossed-looking geometry, there are 10 equations and 9 variables. It is rather complicated. If θ is zero, we get a simplified model, the parallel optical axes model, as shown in Figure 2.3. The epipolar line, is now a horizontal line at the same height as the point pl. Note that $Z_1 = Z_2 = Z$. The depth Z is

$$Z = f - \frac{f\,b}{x_1 - x_2} \tag{2.7}$$

The difference between x_1 and x_2 is known as disparity.

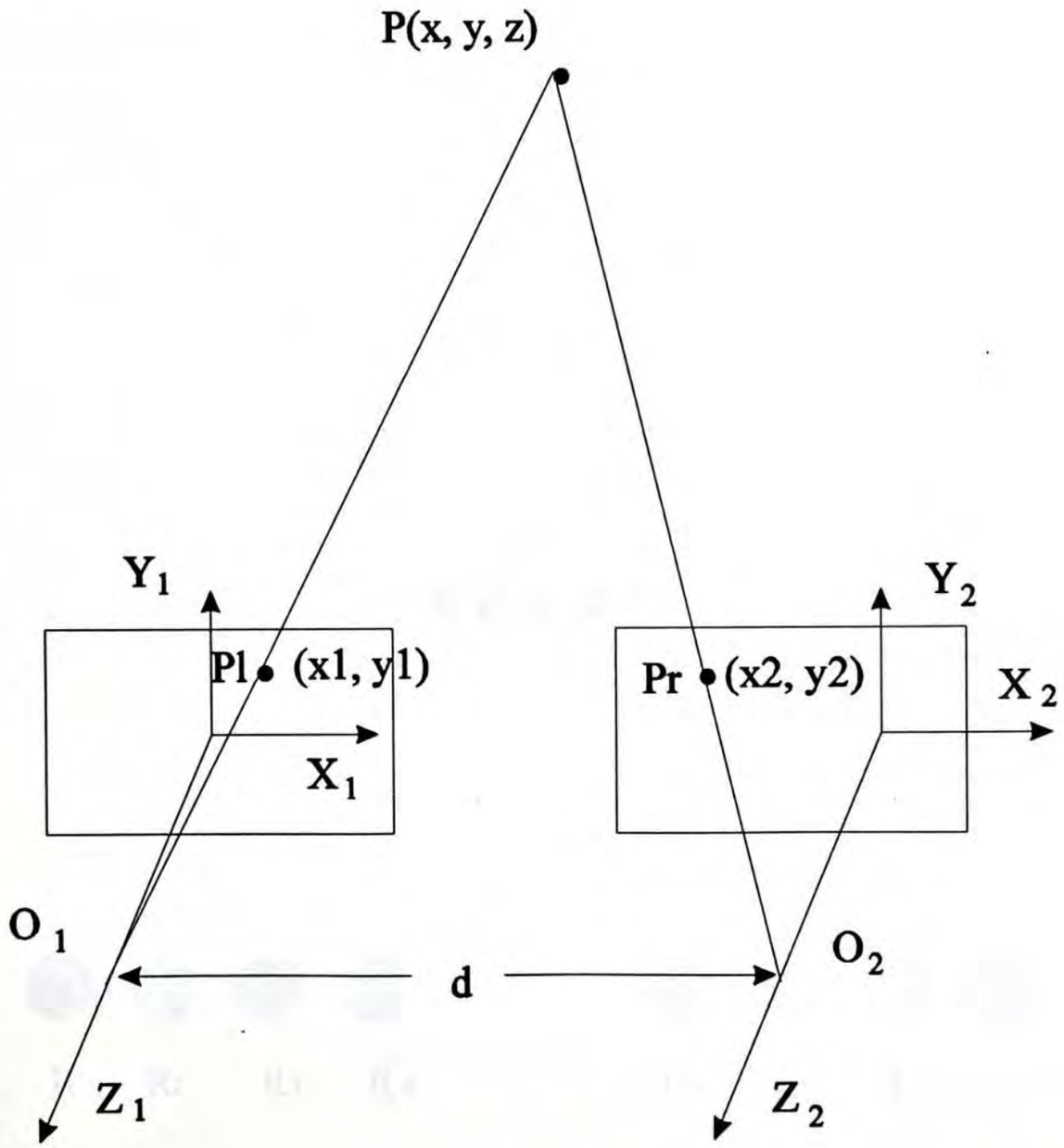


Figure 2.3 The parallel optical axes geometry of stereo vision.

2.2 The false targets problem

If one could identify the exact location of the points, the correspondence problem would be solved easily. However, one cannot mark spots in the scene.

It is very difficult to identify the corresponding locations in the two images because of the false target problem [24]. An example of this problem is shown in Figure 2.4. In this figure, different sets of points in the world can project to the same set of binocular projections. Each of the four projections in one eye's view could match any of the four projections in the other eye's view [24].

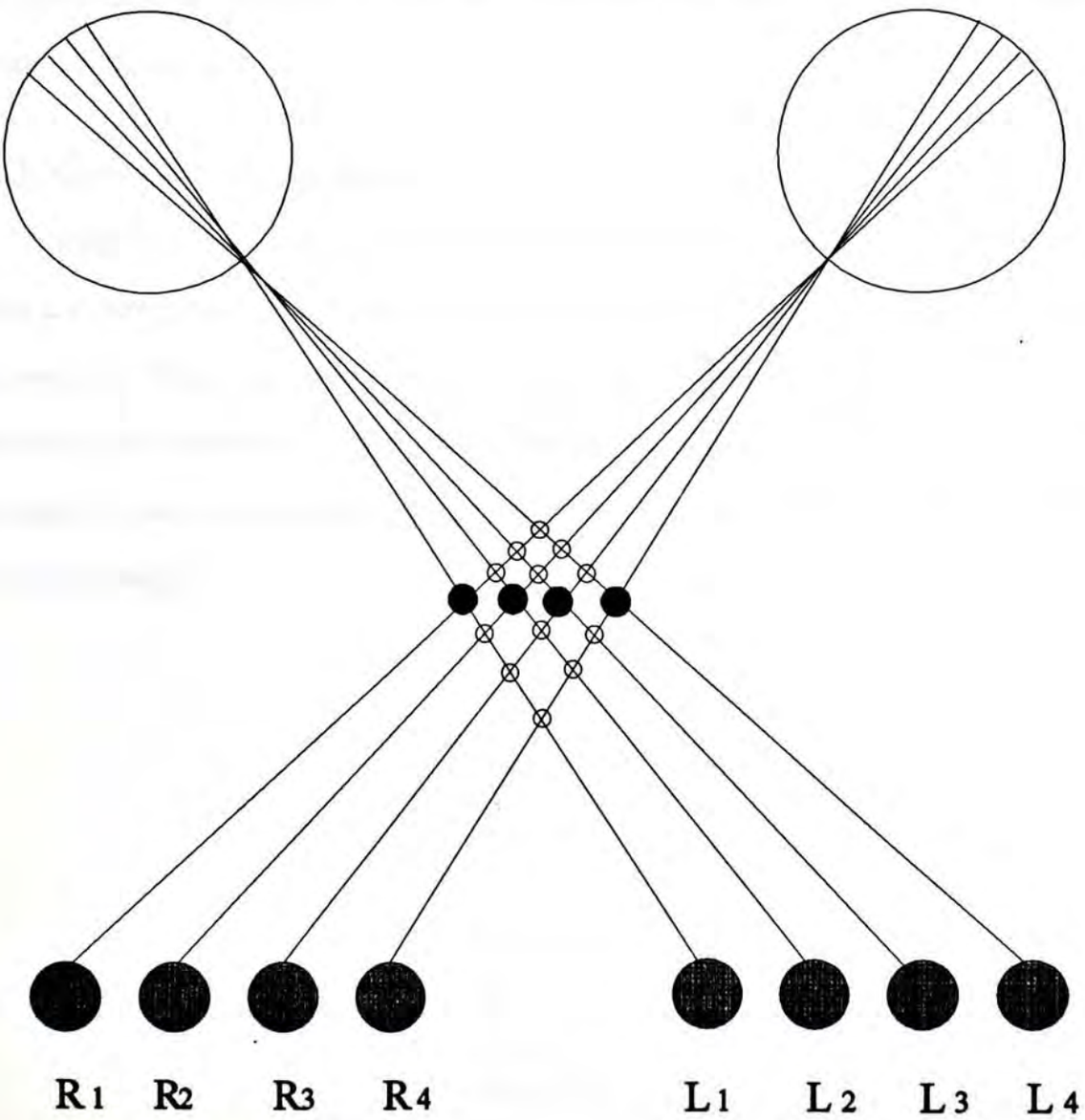


Figure 2.4 The false targets problem in stereo vision. (From Grimson, 1981)

To solve the false target problem, some constraints should be added. They will be discussed in section 2.4.

2.3 Feature selection

Features that are used for the correspondence problem must be reliable. Intensities are unreliable, because two views are differentially distorted due to the photometric and geometric effects. It is not surprising why the gray-level correlation approach for the stereo vision met with limited success. Many physiological evidences and computational analyses of the stereo vision show that edge is the most reliable feature for matching.

2.3.1 Zero-crossing method

Edges are found by looking for places where the gray-level changes abruptly. There are many methods to locate the edges. One is the zero-crossing method that was proposed by Marr and Hildreth [23]. Figure 2.5 shows the first and the second derivatives of intensity. A place where the second derivative crosses zero, or zero-crossing, is where the intensity changes abruptly. We can locate the edges by finding the zero crossings.

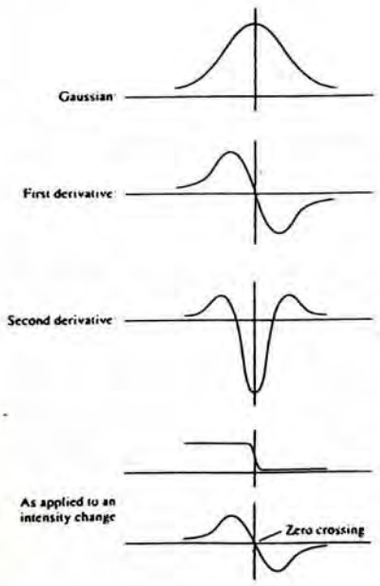


Figure 2.5 The first and the second derivatives of intensity (From Charniak and McDermott, 1985.).

Since the second derivative operator is sensitive to noise, Marr and Hildreth suggested that the best operator for edge detection is the Laplacian of Gaussian ($\nabla^2 G$). The $\nabla^2 G$ is represented by

$$\nabla^2 G = \left(\frac{r^2 - 2\sigma^2}{2\Pi\sigma^6}\right)\exp\left(\frac{-r^2}{2\sigma^2}\right) \tag{2.8}$$

where $r = \sqrt{(x^2 + y^2)}$

The profile of $\nabla^2 G$ is shown in Figure 2.6. The width w in this figure is given by

$$w = 2\sqrt{2}\,\sigma \tag{2.9}$$

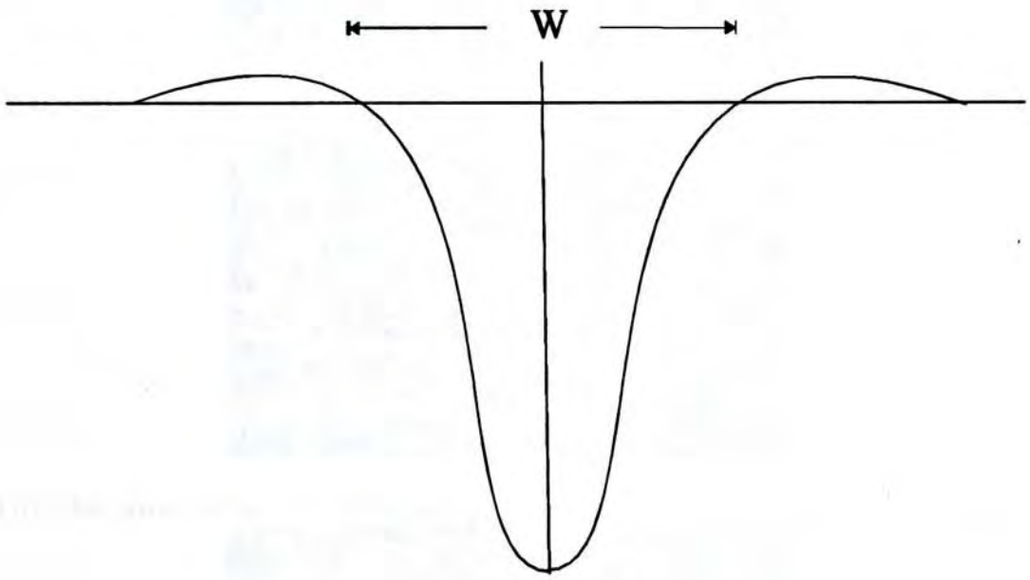
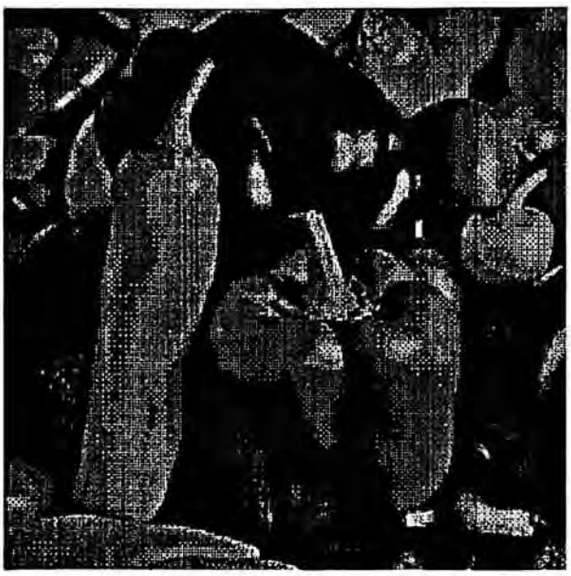
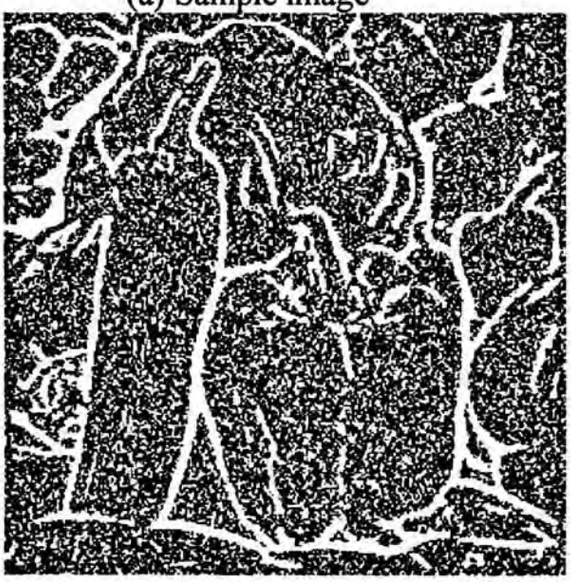


Figure 2.6 The profile of $\nabla^2 G$.

An example of this processing is shown in Figure 2.7.



(a) Sample image



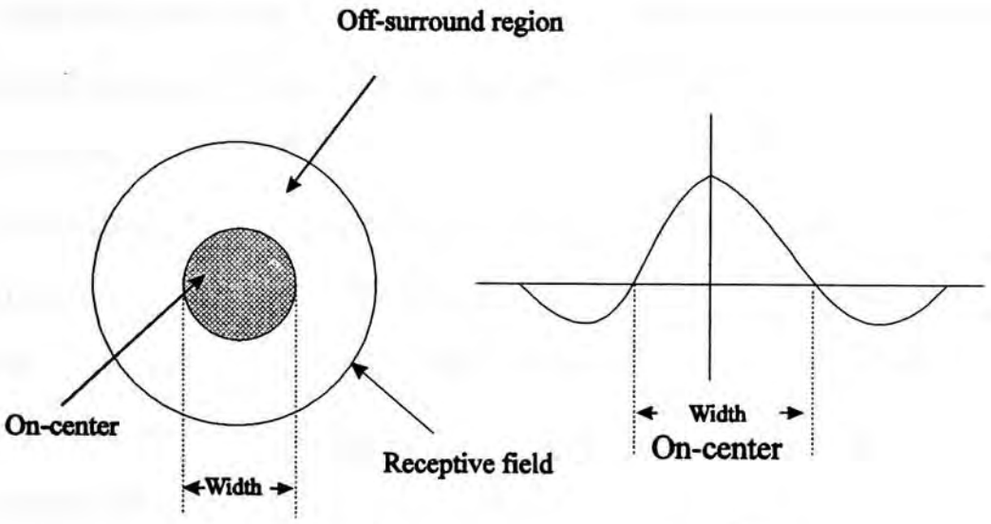
(b) The convolution of the image with the $\nabla^2 G$ operator (σ = 1.5 pixel).

(c) The zero-crossing of (b).

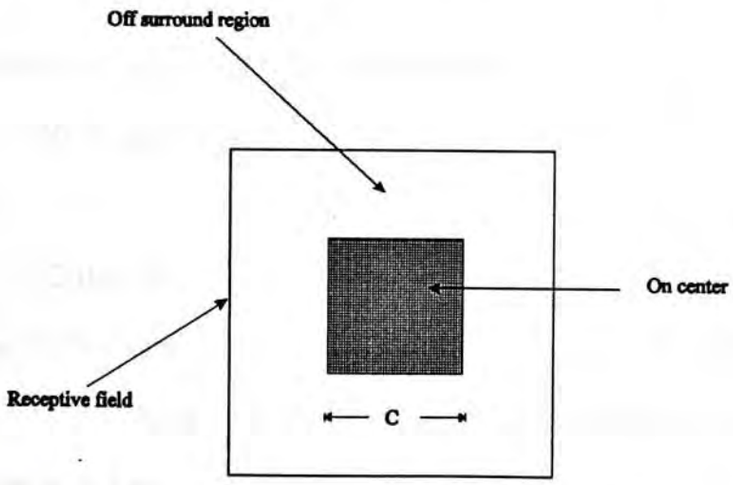
Figure 2.7 Example of using zero-crossing to detect the edges.

2.3.2 A network model for ganglion cell

Marr and Hildreth suggested that the best operator for detecting the edges is the filter $\nabla^2 G$ (Laplacian of Gaussian) and the best approximation of this operator is DOG (difference of two Gaussian Distributions) [23]. Many experiments have shown that the receptive fields of ganglion cells can be described by a DOG [19, 32]. The profile of an on-center receptive field is shown in Figure 2.8 (a). Based on those results, many models have been set up to indicate how visual information is processed [19, 32].



(a) Profile of an on-center receptive field.



(b) Rectangular approximation of a receptive.

Fingure 2.8 An on center receptive field profile and its approximation.

In this chapter, an alternative 3-layer model is proposed. The first layer is the input layer that receives the stimuli from the outside world. The second layer mimics the functions of horizontal bipolar, and amacrine cells in human's retina. The third layer is the output layer that simulates the ganglion cells in which the image information is sent to the higher levels in the visual system for further processing. In this model, each neuron in the output layer has a specific receptive field. The size of the receptive field is determined by the on-center size or off-center size, and the neighboring receptive fields may slightly overlap.

Supposing that there are only two kinds of connections between neurons: one is inhibitory, and the other one is excitatory, then the interconnections of the same level are not considered.

Connections between the layers are local in this model. The connections of an on-type output neuron is shown in Figure 2.9. To set up the relationship between layer 1 and layer 2, the size of the receptive field must be found out first. By approximation, the receptive field is a square (Figure 2.8 (b)) and the relation between the on-center region and the off-surround region is as follows.

Number of neurons in the receptive field (layer 1) is

$$S = (C+2)^2 \tag{2.9}$$

where C is the width of the center region of the receptive field. Number of processing neurons in layer 2 (N) is calculated by the following formula:

$$N = 2*(C+1) \tag{2.10}$$

For example, if $C = 2$ then $S = 16$ and $N = 6$.

If connections between layer 1 and layer 2 are described by a weight matrix W, the relationship between the output of the first layer (I_{ij}) and the neurons in the second layer can be set up as follows.

The first neuron in layer 2 is

$$B_{diagonal\text{-}left} = \sum_{j=0}^{C+1} I_{jj} * W_{jj} \tag{2.11}$$

And the last element of layer 2 is

$$B_{diagonal\text{-}right} = \sum_{i=0}^{C+1} \sum_{j=C+1}^{0} I_{ij} * W_{ij} \tag{2.12}$$

For other neurons, $0 < i$ and $i < C + 1$,

$$B_{horizontal(i)} = \sum_{j=0}^{C+1} I_{ij} * W_{ij} \tag{2.13}$$

and

$$B_{vertical(i)} = \sum_{j=0}^{C+1} I_{ji} * W_{ji} \tag{2.14}$$

where $j = k$. Every weight W_{ij} in the weight matrix W should show its physiological meaning, that is, inhibitory or excitatory connection. Therefore, every element W_{ij} chooses either negative value or positive value that represents inhibitory or excitatory connections respectively. For an on-center receptive field, the weight values are calculated as follows

$$W_{ij} = \begin{cases} \frac{2}{C}, & if\ 0 < i,j \le C \\ -1, & if (i = 0 \vee j = 0) \wedge (C < i,j \le C+1) \end{cases} \tag{2.15}$$

In the same way, we can define relationship between output neuron G in layer 3 and neurons B in layer 2.

$$G = B_{diagonal\text{-}left} + B_{diagonal\text{-}right} + \sum_i B_{horizontal(i)} + \sum_i B_{vertical(i)} \tag{2.16}$$

After output neuron G sums up all the inputs from layer 2, it does a thresholding operation on the result. For an on-center receptive field, its final output gn is

$$gn = \begin{cases} 1, & if\ G > 0 \\ 0, & if\ G \le 0 \end{cases} \tag{2.17}$$

For an off-center receptive field, the weight values are rewritten as follows

$$W_{ij} = \begin{cases} \frac{-2}{C}, & if\ 0 < i,j \le C \\ 1, & if (i = 0 \vee j = 0) \wedge (C < i,j \le C+1) \end{cases} \tag{2.18}$$

and the final output of the off-center neuron (gf) is

$$gf = \begin{cases} 1, & if\ G < 0 \\ 0, & if\ G \geq 0 \end{cases} \tag{2.19}$$

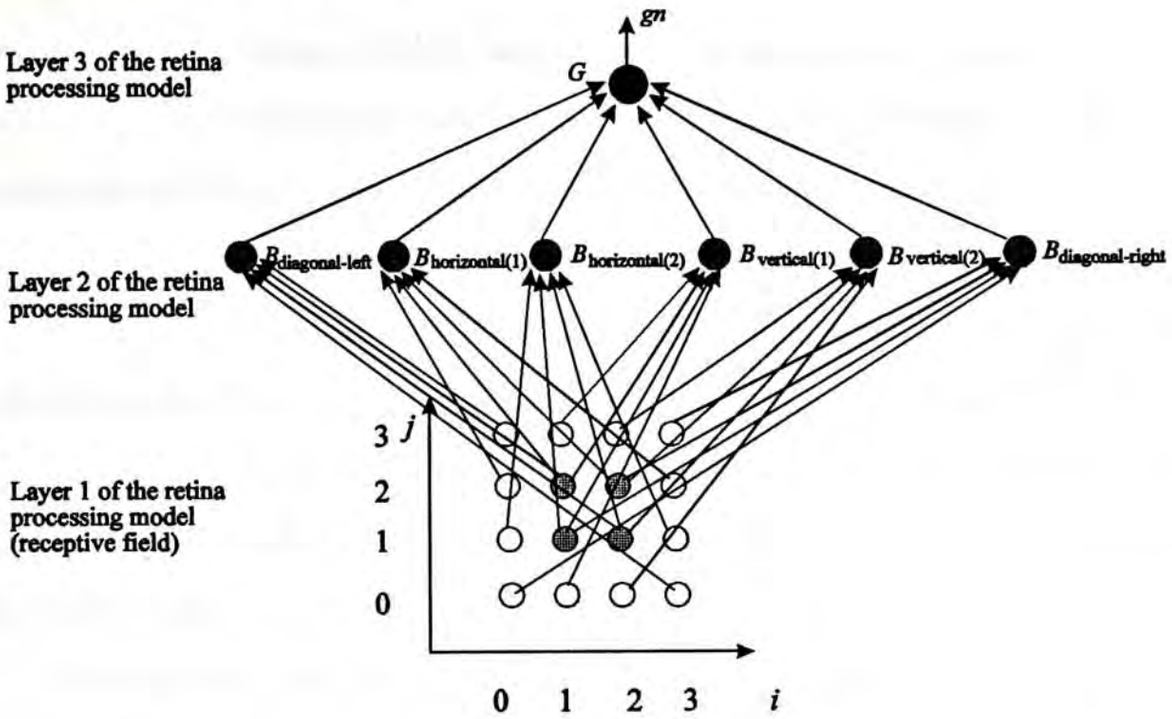
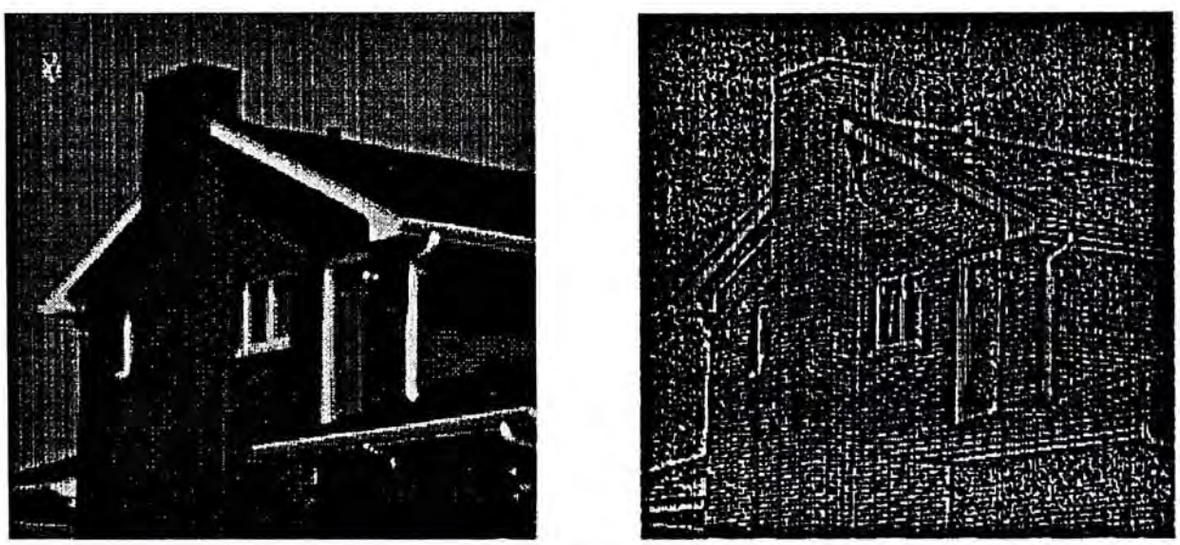


Figure 2.9 Connections of an on-type output neuron in the retina processing model

The convolution of the image with this model is shown in Figure 2.10.



(a) Simple image

(b) The result of the convolution.

Figure 2.10 The convolution of the image with the DOG network model.

2.4 The constraints of matching

Stereo correspondence is the process which matches the features that correspond to the same point in the three-dimensional world from a pair of stereo images. The false target problem, as mentioned in section 2.2, cannot be solved completely even for the use of quite specific matching features. However, it is possible to solve this problem by using constraints. These constraints is derived from the basis of physical properties of the world, and the geometry of the image systems.

The uniqueness constraint is derived from the observation that a given point on a physical surface have a unique position in space at any one time. This implies that an feature in one eye should be matched with only one feature in the other eye [24]. This is a powerful constraint that can reduce a lot of ambiguous matches and was explicitly exploited by many researcher [22, 10, 11,25, 29].

The epipolar constraint is a trigonometric constraint [30]. It was first exploited by Keating and was widely used in stereo matching [2, 10, 11, 25, 29]. The epipolar constraint is always obtained by using the special image geometries, as mentioned in section 2.1. This constraint can reduce the searching space of candidates into one-dimension.

The smoothness constraint is suggested by Marr and Poggio [22]. This is an observation of the physical world. The surfaces of objects are generally smooth in the sense while compared with their distance from a viewer. This translates into the rule that the disparity of the matches varies smoothly almost everywhere over the image [24]. However, this constraint may give wrong matching results along discontinuities in depth [11, 25].

As mentioned above, the smoothness constraints will cause error matching results at sudden changes of depth. Mayhew and Frisby suggested other constraint to adjust this error. They stated that the disparity of the matches varies smoothly along a contour. This continuity constraint is often called figural continuity constraint [25]. It is also widely used in the stereo vision correspondence algorithms.

2.5 Correspondence techniques

There are two types of correspondence techniques in computational stereo vision. One is the area-based corresponding techniques and the other is the feature-based corresponding techniques [15]. The area-based techniques find corresponding points on the basis of similar corresponding areas in the left and right images while the feature-based techniques match features in the left image to those in the right image. The area-based techniques are more robust in the mix of buildings and open terrain scenes. However, the feature-based techniques can provide more accurate information of depth discontinuities and the height of objects [15]. Furthermore, the feature-based techniques are faster than area-based techniques and are less sensitive to photometric variations of the image systems.

2.6 Conclusion of chapter 2

The correspondence problem is a critical problem in the stereo vision. We cannot entirely solve it without any constrain. Some constraints that based on the observation of the physical world was proposed, including, the uniqueness constraint, the epipolar constraint and the continuity constraint. These constraints have been used in designing our correspondence algorithms.

Chapter 3

A Hypercolumn Based Stereo Vision Model

Currently, there are two types of stereo vision models developed by researchers who are interested in the problem of binocular vision. The first one is the computer based stereo vision model and the second one is the psychophysical based stereo vision model. The computer based stereo vision model [1, 2, 11, 15, 20, 22, 28, 29] is a more practial approach. On the other hand, the psychophysical based stereo vision model [10, 12, 13, 22, 25] uses the evidences in visual system of the human to solve the problem. In this chapter, the psychophysical based stereo vision model (PSVM) is proposed. PSVM uses the hypercolumn structure as a topographical mapping of the recptor surface within a three-dimensional structure to achieve the binocular processing.

3.1 A visual model for stereo vision

Since Wheatstone invented the stereoscope in 1835, the processes of stereo vision have been intensively studied. The studies are not only in the areas of psychophysics and physiology, but also in the area of information processing. With evidences from psychophysics and physiology of the human binocular system, a visual model for stereo vision is proposed.

In this model, the visual cortex is columnar organization. It is a very useful structure for computational processing. There are four main columns in that model, including location columns, orientation columns, ocular dominance columns and disparity columns (see Chpater 1) and two type of hypercolumns, hypercolumn 1 and hypercolumn 2. These two hypercolumns are similar to the hypercolumn model from

Hubel and Wiesel [18]. Hypercolumn 1 is formed by the location columns, the orientation columns and the ocular dominance columns (Figure 1.10). Hypercolumn 2 is similar to hypercolumn 1 but the former also includes the disparity columns.

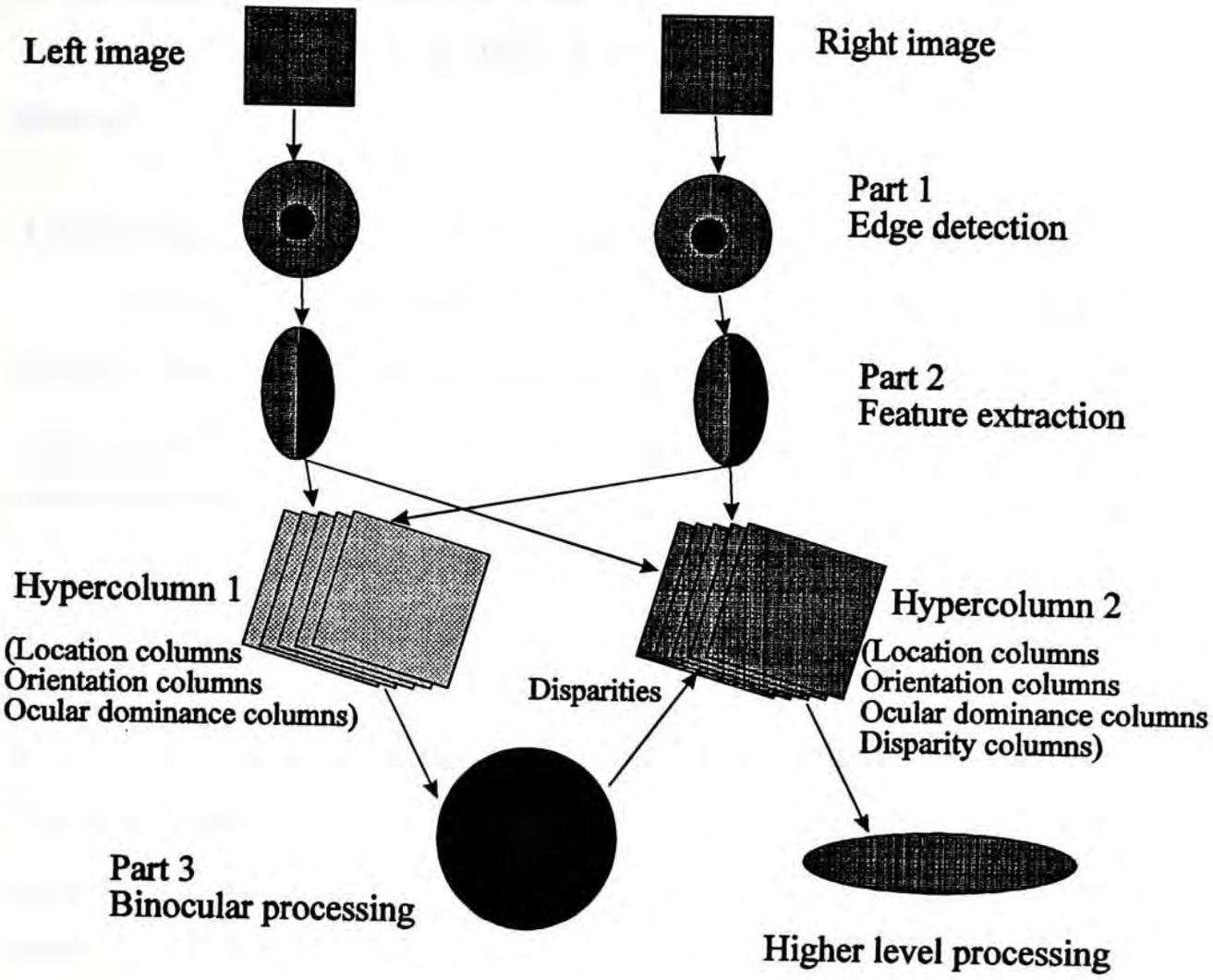


Figure 3.1 A visual model for stereo vision.

The visual model can be functionally divided into three parts: edge detection, feature extraction and binocular processing (Figure 3.1). The edge detection performs the function of ganglion cells. Note that it is possible to use the other edge operators although the function of ganglion cells can be described by DOG. The feature extraction is carry out the functions of simple cells. In this part, the oriented lines will be extracted and placed into the hypercolumns. Note that there are two paralell visual

pathways. One goes to hypercolumn 1 and the other one goes to hypercolumn 2. The binocular processing is to use the information in hypercolumn 1 to solve the stereo correspondence. After this processing, the disparities will be placed into hypercolumn 2. The information in hypercolumn 2 can be used in higher level processing, for example, surface reconstruction. Based on this visual model, the PSVM system is designed.

3.2 The model of PSVM (A Computerized Visual Model)

Figure 3.2 is the block diagram of PSVM. The inputs of PSVM are a pair of left and right images. It consists of three main stages. The first stage is dedicated for local orientation line extraction. The local oriented line extraction includes two steps: edge point detection and orientated line detection. It will extract two types of edges: on-type lines and off-type lines, and four kinds of orientated lines: horizonal lines, vertical lines, left diagonal lines and right diagonal lines. At the end this stage, lines of different orientations are placed into the hypercolumns (hypercolumn 1 and hypercolumn 2). The local orientated line extraction will be discussed in section 3.3. The second stage is used for local lines matching where orientation lines in hypercolumn 1 are matched and the local disparities are found. The local lines matching comprises two kinds of matchings, short orientated lines matching and long oriented lines matching. They can be done in parallel. Stage 2 will be described in section 3.4. The third stage is for the disparity integrations. Local disparities are integrated and the final disparities (global disparities) are placed into hypercolumn 2 for higher level processing at this stage. The disparity integrations uses competition model to select the right disparity in a specific area. Once the right disparity is selected, it will be reassigned to this area. Stage 3 will be discussed in section 3.5.

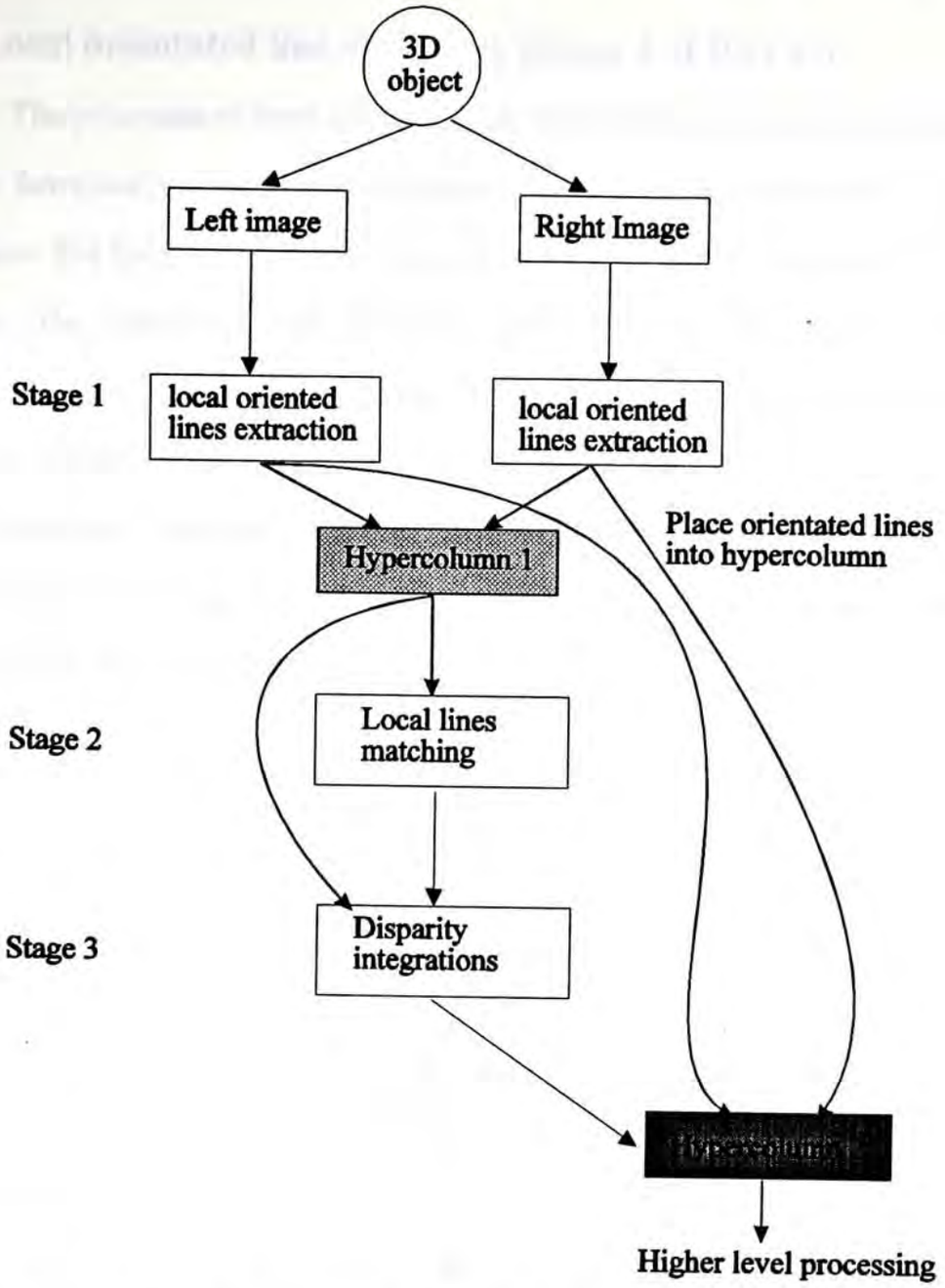
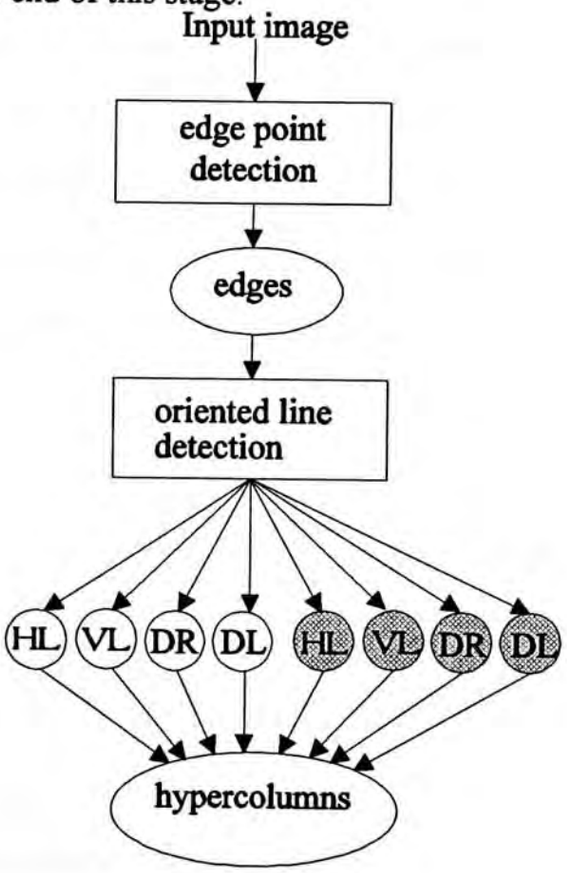


Figure 3.2 The block diagram of PSVM (a computerise visual model for stereo matching). There are three stages in PSVM. After stage 1, the features (oriented lines) of images will be extracted and stored into the hypercolumn 1 and hypercolumn 2. This stage is called local oriented lines extraction and will be discussed in section 3.3. Stage 2 is local lines matching. This stage uses the information in hypercolumn 1 to achieve the matching (see section 3.4 for details). Local disparities produced by stage 2 will be integrated at stage 3 (disparity integrations). Note that the information in hypercolumn 1 is also used by the disparity integrations to get disambigous disparities (see section 3.5 for details). The results of the stage 3 will be stored into hypercolumn 2 for higher level processing.

3.3 Local orientated line extraction (Stage 1 of PSVM)

The processes of local orientated line extraction are presented in Figure 3.3. It can be functionally divided into two steps: edge point detection and orientated line detection. The function of edge point detection is to detect the edge points in the input images. The orientated line detection arranges these edge points so that the orientations of the lines can be found. The local orientated line extraction model can detect two types of edges: the on-type lines and the off-type lines, and four kinds of directional lines: horizonal lines, vertical lines, left diagonal lines and right diagonal lines. These orientation lines are placed into the hypercolumns (hypercolumn 1 and hypercolumn 2) at the end of this stage.



HL Horizontal line
VL Vertical line
DR Right diagonal line
DL Left diagonal line

On-type line

Off-type line

Figure 3.3. The processes of local orientated line extraction (Stage 1 of PSVM)

3.3.1 Orientated line detection network

The orientated line detection network may be considered as the simple cells described by Hubel and Wiesel [16, 18]. In Hubel and Wiesel's model, the receptive fields of lateral geniculate cells are arranged along a straight line on the simple cells (Figure 3.4). Marr and Hildreth also suggested a mechanism for detecting orientated zero-crossing segments [24]. The idea is that an on-center receptive field signal corresponds to a peak positive value of the filtered image $\nabla^2 G * I$, and an off-center receptive field signal corresponds to the contrary, a peak negative values. The sum of their firings will correspond to the slope of the zero-crossing so that an edge must pass between on-center and off-center receptive fields (Figure 3.5).

In the current implementation, the orientated line detection network is a 2-layer network (Figure 3.6). The first layer is the input layer that takes the output from the edge point detection network, and the second one is the output layer. Four orientated lines, horizontal, vertical, left diagonal and right diagonal lines, can be detected by this network. The patterns of these four detectors are shown in Figure 3.7.

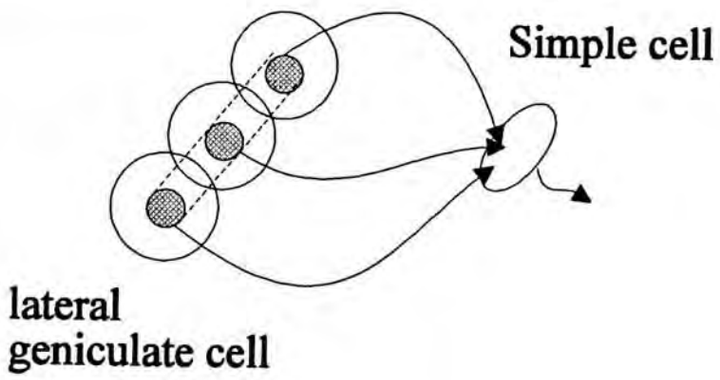


Figure 3.4 The model of Hubel and Wiesel for explaining the organization of simple cell.

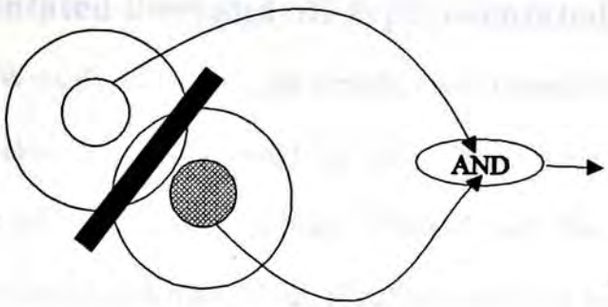


Figure 3.5 The model of Marr and Hildreth for detecting orientated zero-crossing segments.

The outputs of the orientation line detection network can be described by

$$S_k = \sum_i gn_i + \sum_j gf_j \tag{3.1}$$

where *gn* is the on-center receptive field signal and *gf* is the off-center receptive field signal in the edge point detection network, and,

$$O_k = \begin{cases} 1, & if\ S_k = THD \\ 0, & if\ S_k \neq THD \end{cases} \tag{3.2}$$

where THD is a threshold.

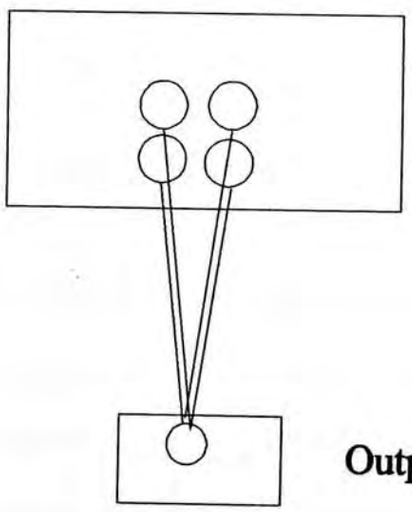


Figure 3.6 The connections of orientated line detection network

3.3.2 On-type orientated lines and off-type orientated lines

In Hubel and Wiesel's model of the simple cells (mentioned in section 3.3.1), on-center receptive fields can be replaced by off-center receptive fields then these simple cells can detect off-type orientated lines. These means that the on-center simple cell and the off-center simple cell may occur simutanlously and have similar functions. In this model, these two types of simple cell are simulated.

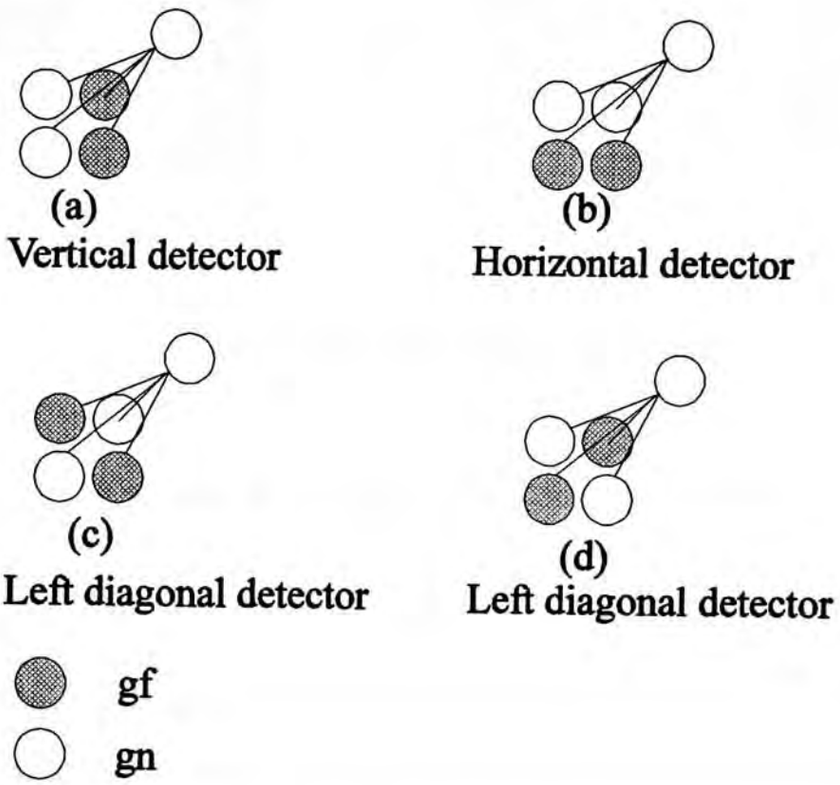


Figure 3.7 Four types of detectors in orientated line detection network

In the orientated line detection network (described in section 3.3.1), on-type orientated lines and off-type orientated lines can be detected. The on-type orientated lines can only be detected by the on-center simple cells and the off-type orientated lines can only be detected by the off-center simple cells. The patterns of the on-type orientation lines and the off-type orientation lines are shown in Figure 3.8. By comparing Figure 3.8 patterns with the ones in Figure 3.7, we know that the orientation line detection network are composed of on-center simple cells.

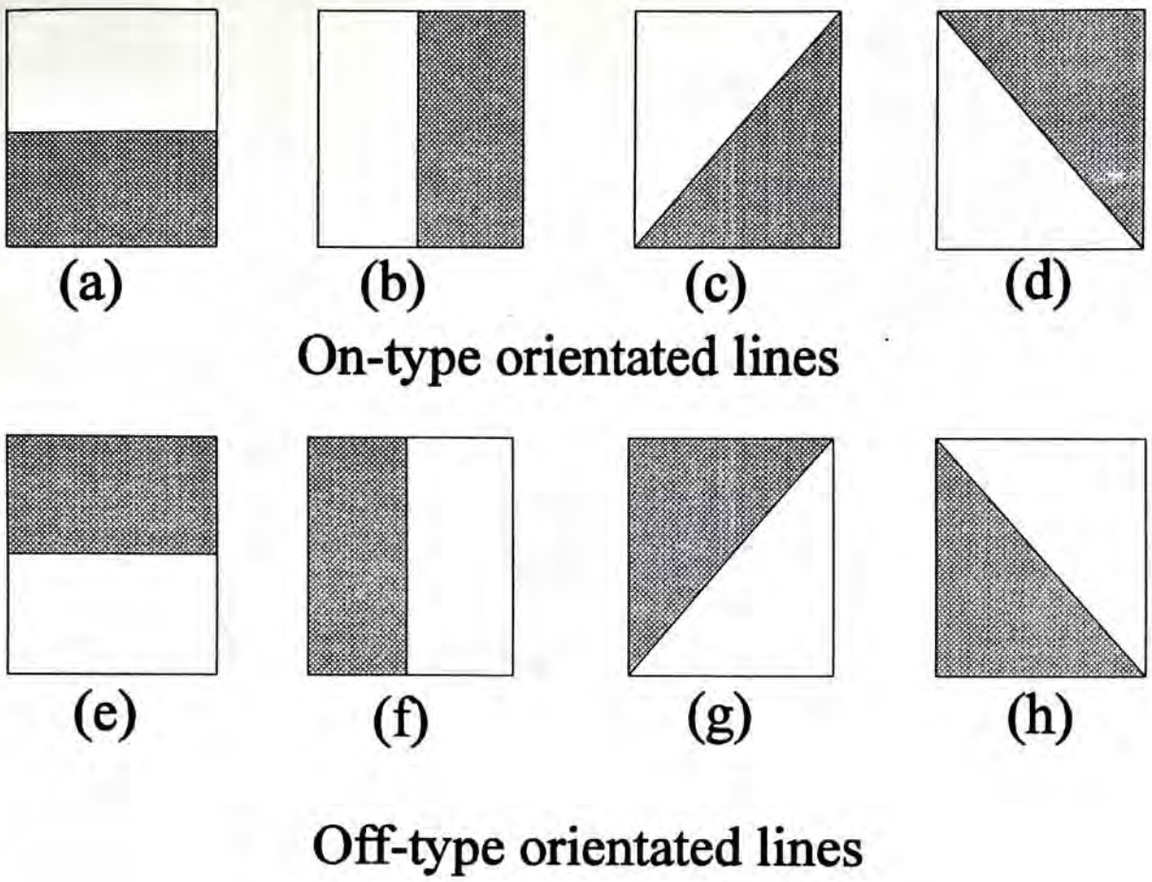


Figure 3.8 The patterns of on-type and off-type orientatied lines in PSVM.

3.4 Local line matching (Stage 2 of PSVM)

The local line matching processes are shown in Figure 3.9. It contains two main models, including the line length discrimination model and matching model. Orientated lines come from the local line extraction layer are sorted and placed in the hypercolumns and are ready for matching. There are two types of matchings, including short orientated lines matching and long orientated lines matching. They can be performed in parallel. The line length discrimination model is used to construct the long orientated lines (see section 3.4.2 for details). The matching model is used to find out local disparities and will be discribed in section 3.4.5.

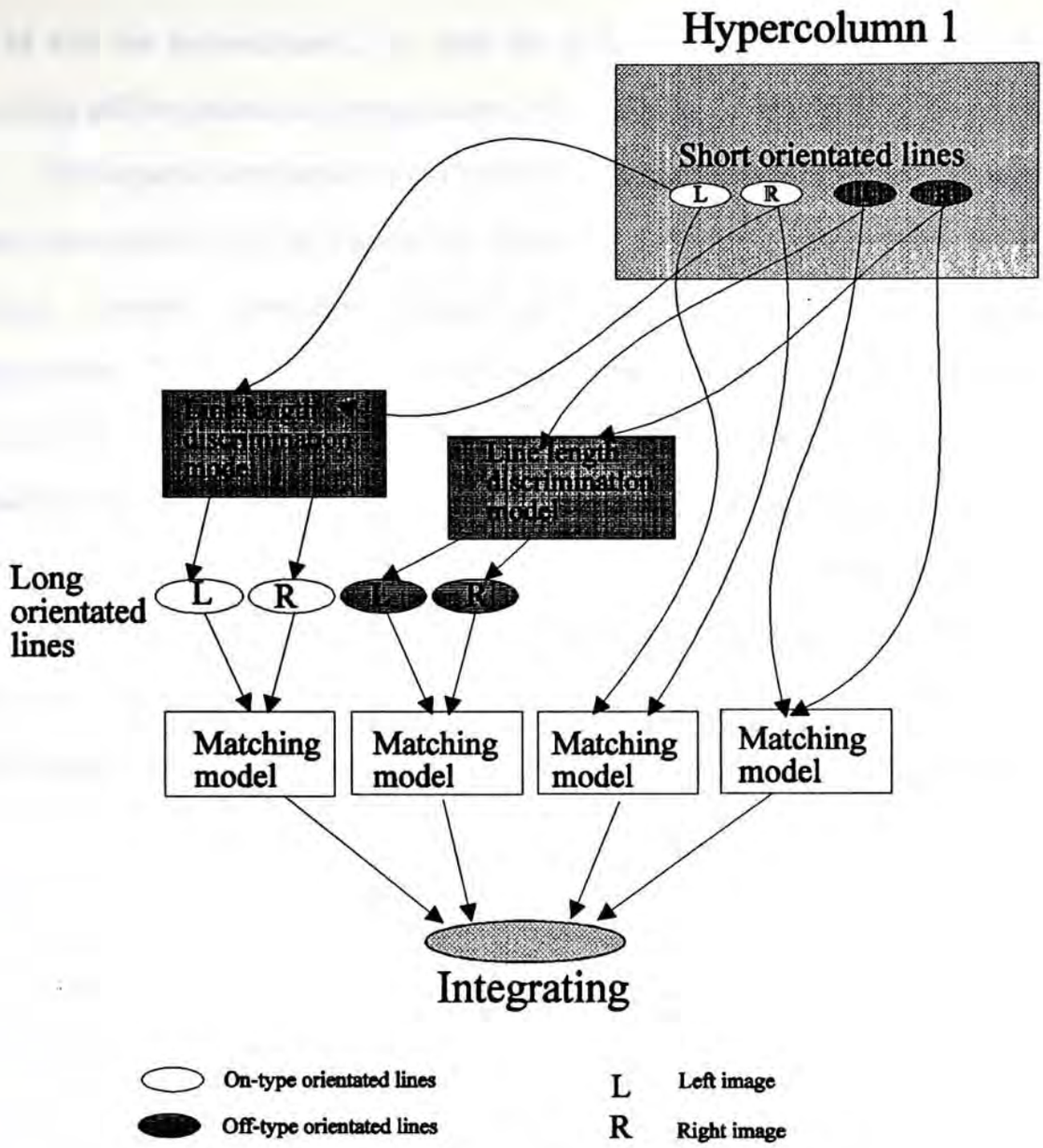


Figure 3.9 Block diagram of local line matching in PSVM. Line length discrimination model is used to construct long oriented lines for long orientated lines matching and is described in section 3.4.2. The matching model is used to find out local disparities and is illustrated in section 3.4.5.

3.4.1 Structure of hypercolumn in PSVM

As mentioned in section 3.1, two types of hypercolumns, hypercolumn 1 and hypercolumn 2, are included in PSVM. The structure of these two hypercolumns are similar. The only difference is that the hypercolumn 2 has the disparity column while the hypercolumn 1 has not. The hypercolumn 1 is used for matching in the stage 2 of

PSVM and the hypercolumn 2 is used for higher level processing. Note that the disparities will be placed into hypercolumn 2 after the stage 3 of PSVM.

The hypercolumn structure in PSVM is similar to the one in Hubel and Wiesel's model (see section 1.5.2 in Chapter 1). There are three kinds of columns, including location columns, orientation columns and ocular dominance columns in a hypercolumn. The shape of the hypercolumn is like a cube. Figure 3.10 shows the structure of the hypercolumn in PSVM. The location column is organized as a Euclidean plane which is a natural representation of images. The orientation column is perpendicular to this plane while the ocular dominance columns are rearranged so that the binocular processing is clear. In Figure 3.10, the left ocular dominance columns are marked by L and the right ocular dominance columns are marked by R. The structure of left hypercolumn (L) elements and right hypercolumn (R) elements are the same.

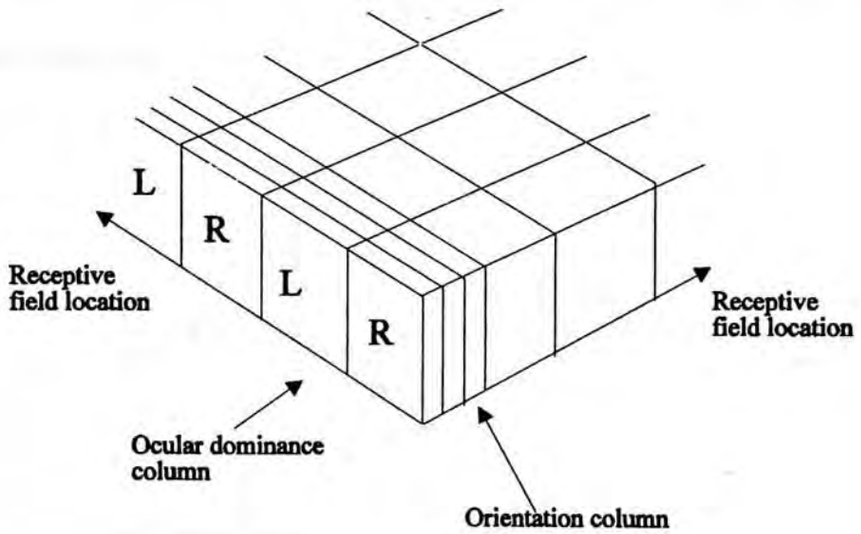


Figure 3.10 The hypercolumn structure in PSVM

A element of hypercolumn 1 is defined as follows.

$$h_{tij} = \begin{cases} 1, & \textit{if the oriented line of h is t} \\ 0, & \textit{otherwise} \end{cases} \tag{3.3}$$

where i, j is the position of the hypercolumn element and t is the types number of orientation. The menaings of t are given by

$$t = \begin{cases} 0, & \textit{horizontal} \\ 1, & \textit{vertical} \\ 2, & \textit{left diagonal} \\ 3, & \textit{right diagonal} \end{cases} \tag{3.4}$$

The representation of hypercolumn 2 elements is same as hypercolumn 1 but the value of hypercolumn 2 is disparity.

3.4.2 Line length discrimination model (Part of stage 2 of PSVM)

As mentioned above, there are two types of matchings, short orientated lines matching and long orientated lines matching. Their matching mechanisms are identical. However, line discrimination is required before the long orientated lines matching can start (Figure 3.9). The line discrimination mechanism is to reconstruct short orientated lines into a long one. This mechanism might be considered as a neural mechanism of line length discrimination.

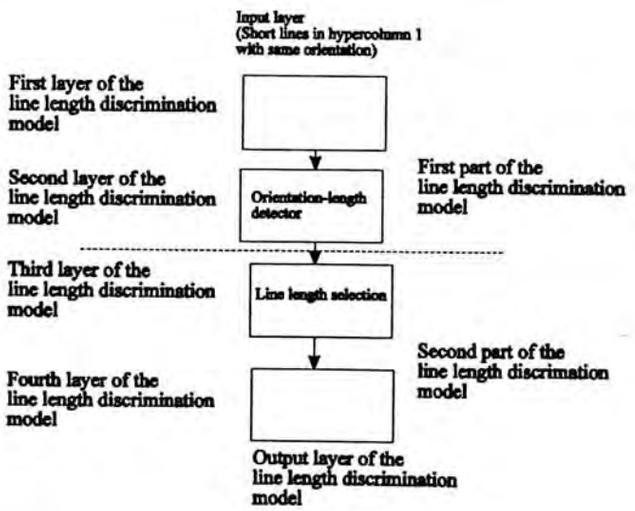


Figure 3.11 The line length discrimination model (part of stage 2 of PSVM). This is a four-layer model. The first layer is input layer. The second layer is orientation-length detectors which detect the length of specific oriented lines (see section 3.4.3). The third layer is line length selection layer which selects the longest line from differnt orientation-length detectors (see section 3.4.4). The result is outputted by the fourth layer.

Figure 3.11 depicts the line length discrimination model. This is a four-layer model which has two parts. The first part is for line reconstruction and the second part is for length discrimination. The first part consists of the first layer and the second layer while the second part comprises the other two layers. The first layer is the input layer which recieves signals from the same orientation hypercolumns (hypercolumn 1). The second layer and the third layer are intermediate layers. Cells in the second layer are orientation-length detectors. Each cell has its preferred orientation and preferred length of lines. Note that a long oriented line may excite a group of orientation-length detectors. The third layer (line length selection layer) is used to select the longest line and suppresses the other. The length of a line is outputted by the fourth layer.

3.4.3 Orientation-length detector

The function of the orientation-length detector is similar to that of the orientated line detector (mentioned in section 3.3.1). However, the orientation-length detector has various sizes and detects orientated lines roughly. A orientation-length detector of specific size can only response to a specific line that has specific length and rough orientation. In PSVM, four types of orientation-length detectors are simulated. Figure 3.12 shows the patterns of these four types of orientation-length detectors. The connection of the detectors are described by the formula

$$I_{t,s} = \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} h_{tij} * W_{tsij} \qquad (3.5)$$

where s is the size of the detector, t is integer representing the orientation of the line. h is the element of hypercolumn 1 and W is the weight matrix of the corresponding detectors. The value of t is between 0 and 3 corresponding to the orientation of horizontal, vertical, left diagonal and right diagonal. The weight matrixes are as follows.

For horizontal lines detection, the weight matrix is

$$W_{0sij} = \begin{cases} 1, & if\ (0 \le i \le s-1) \wedge (0 < [\frac{s}{2}] - 1 \le j \le [\frac{s}{2}]) \\ -1, & otherwise \end{cases} \quad (3.6)$$

For vertical lines detection, the weight matrix is given by

$$W_{1sij} = \begin{cases} 1, & if\ (0 \le j \le s-1) \wedge (0 < [\frac{s}{2}] - 1 \le i \le [\frac{s}{2}]) \\ -1, & otherwise \end{cases} \quad (3.7)$$

For left diagonal lines, the weight matrix is defined by

$$W_{2sij} = \begin{cases} 1, & if\ (0 \le i, j \le s-1) \wedge (i = j \vee i = j \wedge 0 \le i < j+2) \\ -1, & otherwise \end{cases} \quad (3.8)$$

For right diagonal lines, the weight matrix is

$$W_{3sij} = \begin{cases} 1, & if\ (0 \le i, j \le s-1) \wedge (i+j = s-1 \vee i+j = s-1 \wedge 0 \le i-1 \le j \le s-1) \\ -1, & otherwise \end{cases} \quad (3.9)$$

The definition of the second layer 's output is

$$IO_{t,s} = \begin{cases} 1, & if\ I_{t,s} > threshold \\ 0, & otherwise \end{cases} \quad (3.10)$$

For example, if $IO_{0,4} = 1$ then a horizontal line, the length is 4 units, is detected.

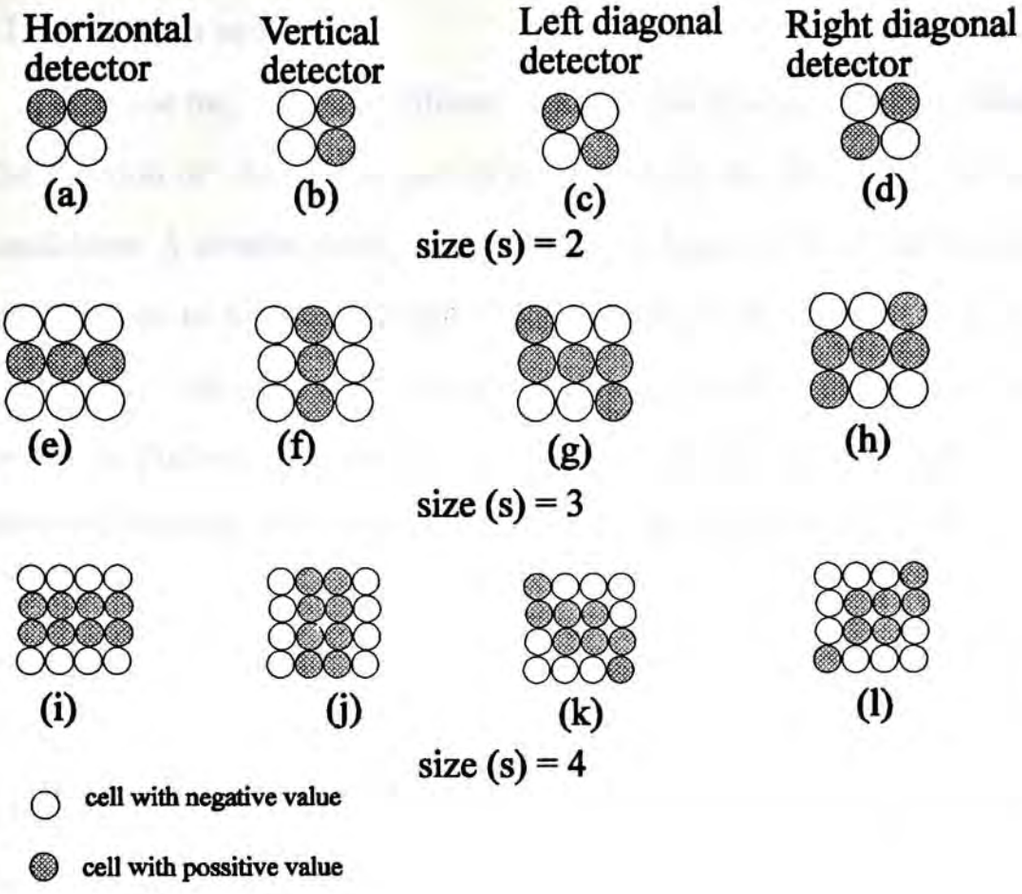


Figure 3.12 Orientation-length detectors in the line length discrimination model (four different sizes of the orientation-length detectors are shown). Note that the length of a orientated line is equal to the size of the detector s.

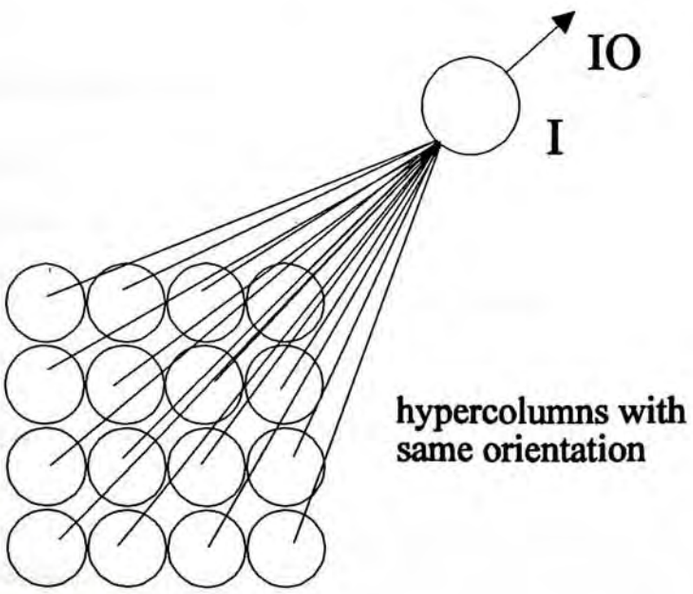


Figure 3.13 The connections of the orientation and line length detector.

3.4.4 Line length selection

A long line may trigger a different group of the orientation-length detectors to fire. The function of the second part of the line-length discriminator is to select the right candidates. A detailed model of it is shown in Figure 3.14. A cell in the second layer corresponds to a specific length of the line (M_0 is the shortest and M_n is the longest). If more cells are added in the second layer, more lines of different length can be detected. In Figure 3.14, there are two kinds of cells, the starting cell (S cell) and the middle cell (M cell). The connections of these cells are slightly different.

For S cells,

$$M_0 = IO_{t,s=1} \tag{3.11}$$

$$L_n = M_n \tag{3.12}$$

where t, s and $IO_{t,S}$ are defined in section 3.4.3, and the corresponding outputs of the S cells are

$$MO_0 = \begin{cases} 1, & if\ M_0 \geq 1 \\ 0, & otherwise \end{cases} \tag{3.13}$$

$$LO_n = \begin{cases} 1, & if\ L_n \geq 1 \\ 0, & otherwise \end{cases} \tag{3.14}$$

For M cells,

$$M_i = M_{i-1} + IO_{t,s=i} \tag{3.15}$$

where $i = 1..n$, and the output is

$$MO_i = \begin{cases} 1, & if\ M_i \geq 1 \\ 0, & otherwise \end{cases} \tag{3.16}$$

and the connections of the M cells in the fourth layer are

$$L_i = MO_i - MO_{i+1} \tag{3.17}$$

where $i = 0..n\text{-}1$, and the output of L_i is

$$LO_i = \begin{cases} 1, & if\ L_i \geq 1 \\ 0, & otherwise \end{cases} \tag{3.18}$$

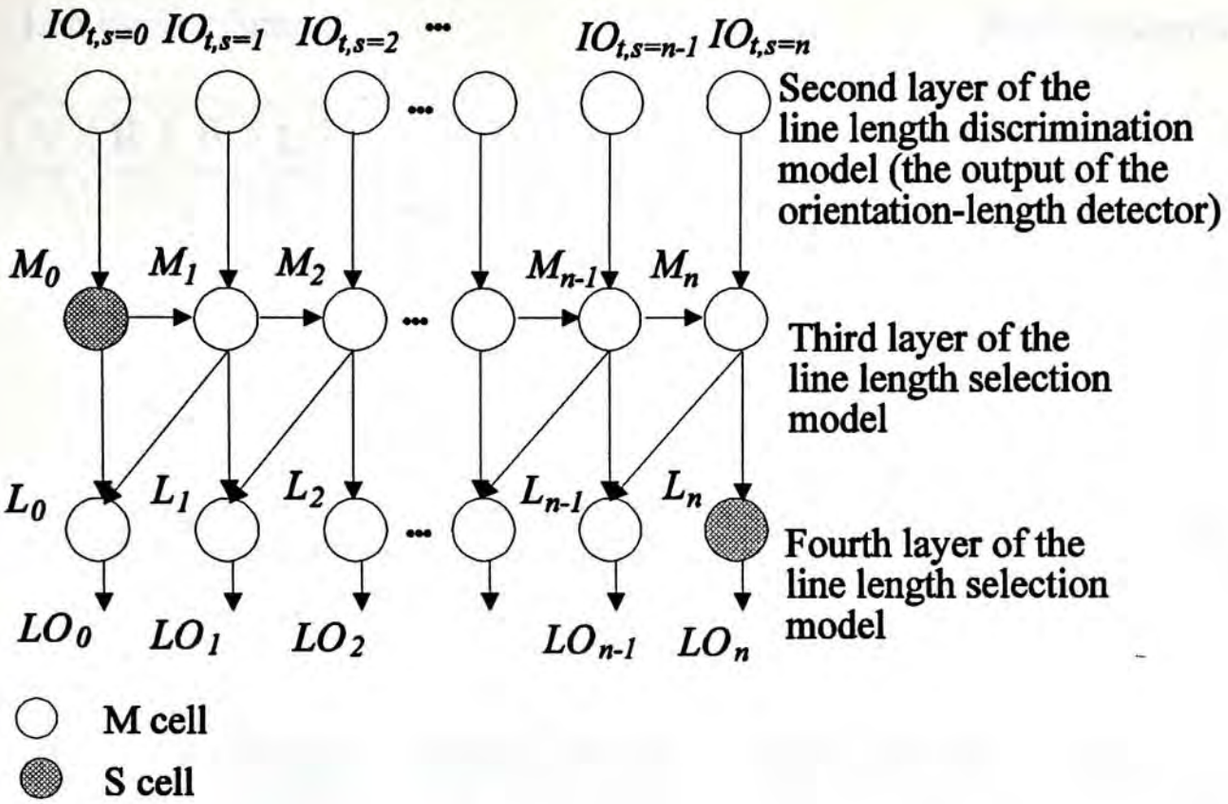
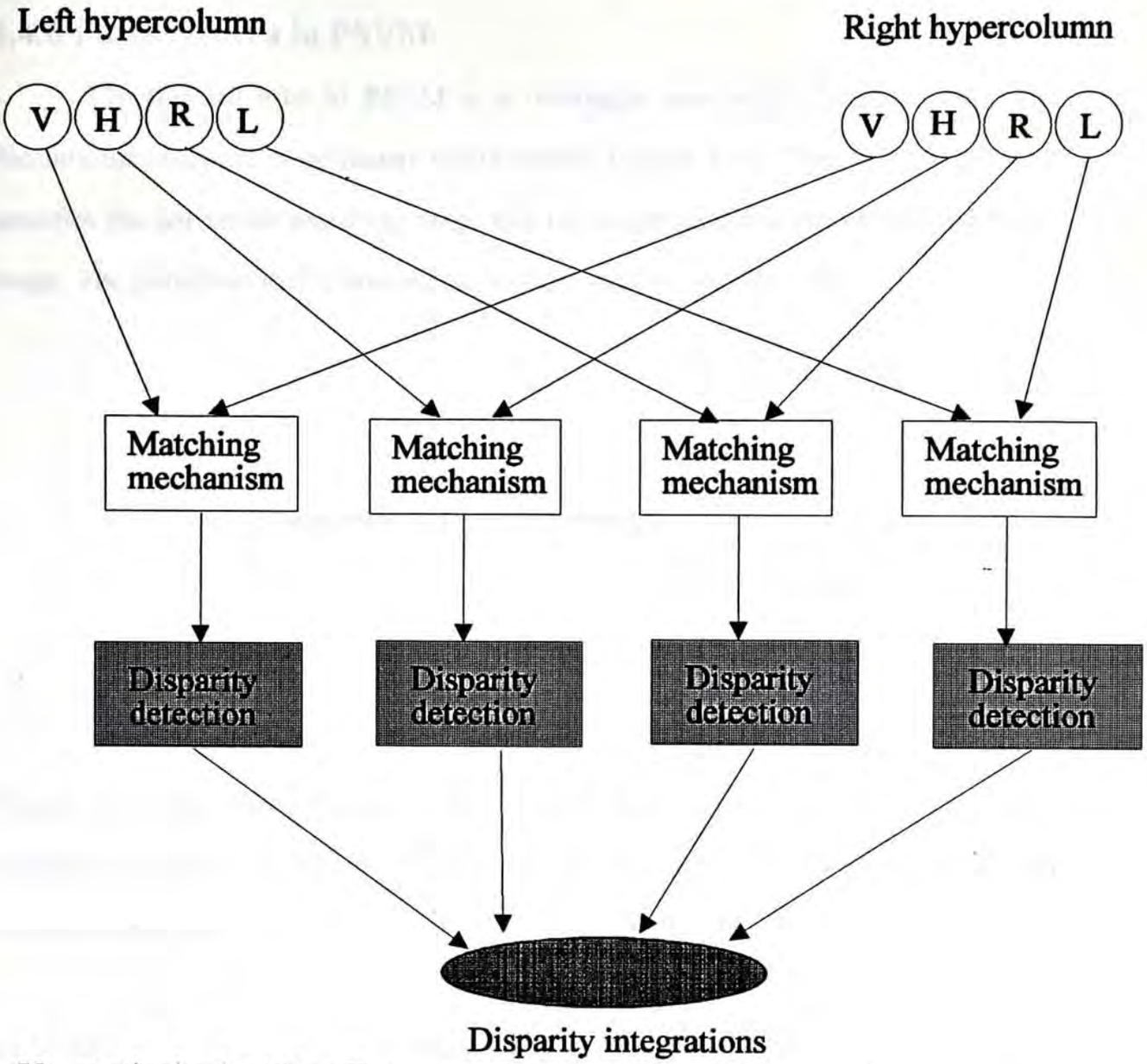


Figure 3.14 The line length selection model

3.4.5 The matching model

The matching model is illustrated in Figure 3.15. Local lines matching in PSVM are in parallel. There is no interaction between different kinds of orientation hypercolumns. As the matching of lines takes place at the same orientation hypercolumns, horizontal lines in the left image only match horizontal lines in the right image and left vertical lines only match right vertical lines, etc. Matching in PSVM is not restricted to one dimension. It matches possible primitives over an area that is called the fusional area.



V vertical orientation line
H Horizontal orientation line
R right diagonal orientation line
L left diagonal orientation line

Figure 3.15 Block diagram of the matching model in PSVM. Matching in PSVM are in parallel. There is no interaction between different kinds of orientation hypercolumns. The matching mechanism will be discussed at section 3.4.7. After the matching, local disparities will be detected by the process disparity detection (see section 3.4.8). At the disparity integrations (see section 3.5), that local disparities will be integrated and then the final disparities will be found.

3.4.6 Fusional area in PSVM

The fusional area in PSVM is a rectanglar area which corresponds to the Panum's fusional area in the human vision system (Figure 3.16). The width of this area specifies the horizontal searching range and the height specifies the vertical searching range. The primitives in this area are possible candidates for matching.

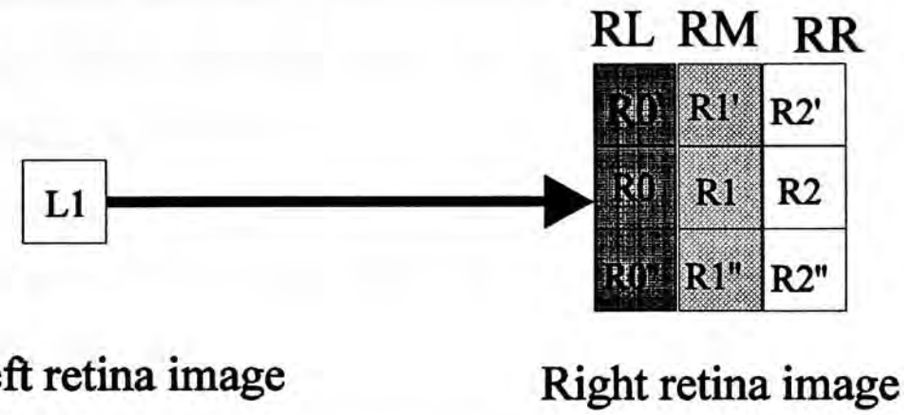


Figure 3.16 The fusional area in PSVM. The fusional area is approximate by a rectangle in PSVM. A left retina image may match a group of right retina images in that fusional area.

In PSVM, the fusional area contains eight hypercolumn elements (Figue 3.16). R1 is the corresponding hypercolumn element of L1 (the hypercolumn element of the left retina image) in the right retina image. R0 is the left hypercolumn element of the R1 and the right hypercolumn element of R1 is R2. The corresponded upper elements of R0, R1 and R2 are R0', R1' and R2', while R0", R1" and R2" are the lower elements. A hypercolumn element of the left image corresponds to a matching set of hypercolumn elements in the right image. For example, in Figure 3.16, a complete matching set of L1 is R0, R0', R0", R1, R1', R1", R2, R2' and R2". This matching set can be divided into three groups, RL, RM and RR, respectively. RL is on the left of L1 and its elements are R0, R0' and R0". Group members of RM are R1, R1' and R1" which are the corresponding points of L1. On the right of L1 is RR whose members are R2, R2', and R2". Note that the width of RM is narrow than RL and RR, and the width of

RL equals to RR.

3.4.7 Matching mechanism

There are two types of matching in PSVM, on-type matching and off-type matching. The on-type matching responds to the on-type orientation lines while the off-type matching processes the off-type orientation lines. The off-type matching is the complement of on-type matching. Here, only the on-type matching mechanism is discussed in detail. In on-type matching, two neurons are wired together only if they have the same orientation and the same length. In PSVM, the neurons are wired between the same kinds of hypercolumns but their ocular dominance columns are different. Figure 3.18 describes how a left retina image (L) at location x matches a right retina image. The matching can be further divided into the L-RL, the L-RM and the L-RR matchings. RL, RM and RR are illustrated in section 3.4.6. The L-RR matching is similar to the L-RL matching except that the searching direction is opposite to each other. The connections of these neurons can be formulated as

$$H_i = \begin{cases} IO_{t,s}, & if\ C_i = 0 \\ 0, & if\ C_i = 1 \end{cases} \tag{3.19}$$

where $IO_{t,S}$ is defined in section 3.4.3 and C_i is as

$$C_i = \begin{cases} 1, & if\ (r_t = 1) \wedge (N_i = 1) \\ 0, & otherwise \end{cases} \tag{3.20}$$

where $t = 0..2$, r_t is the feedback signal will be defined in section 3.4.8 and N_i is

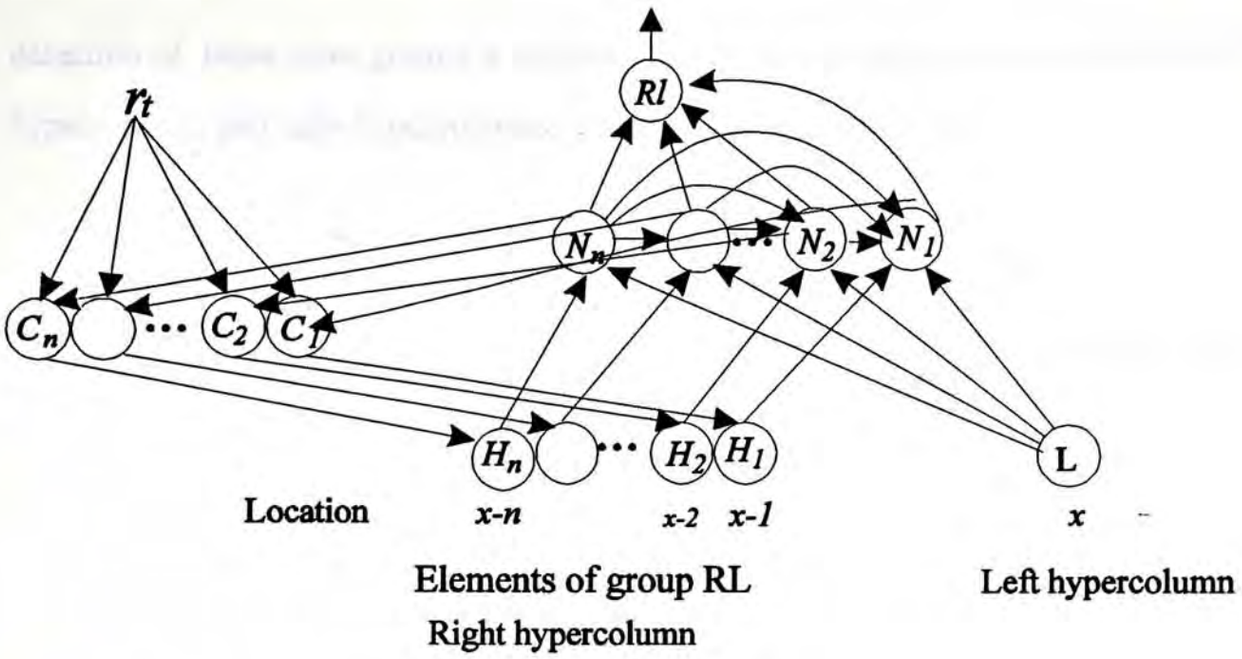
$$N_i = \begin{cases} 1, & if\ [(\sum_{j=i-1}^{n} N_j) = 0] \wedge (L > 0) \wedge (H_i = 1) \\ 0, & otherwise \end{cases} \tag{3.21}$$

where n is the width of the hypercolumn and $i = 1 .. n$

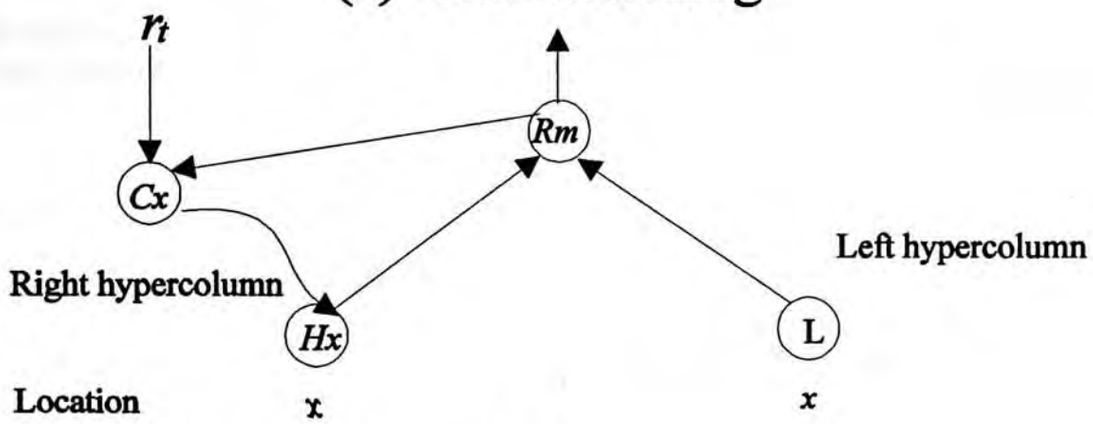
$$RI = \sum_{i=1}^{n} N_i * i \tag{3.22}$$

$$Rm = \begin{cases} 1, & if\ (C_i > 0) \wedge (L > 0) \\ 0, & otherwise \end{cases} \tag{3.23}$$

The representation of *Rr* is the same as *Rl*.



(a) L-RL matching



(b) L-RM matching

Figure 3.18 Matching mechanism in PSVM.

Note that the neuron C_i is the control neuron which will supress the hypercolumn element H_i if it is fired (see section 4.5.2 in Chapter 4 for details).

3.4.8 Disparity detection

In PSVM, disparities can be classified into the near, zero and far disparity

groups. Disparities in these groups are detected by three kinds of neurons, namely, the near, the zero, and the far neurons. As mentioned above, in PSVM, the disparity detection of these three groups is accomplished by wiring depth neurons between left hypercolumns and right hypercolumns. This is shown in Figure 3.19.

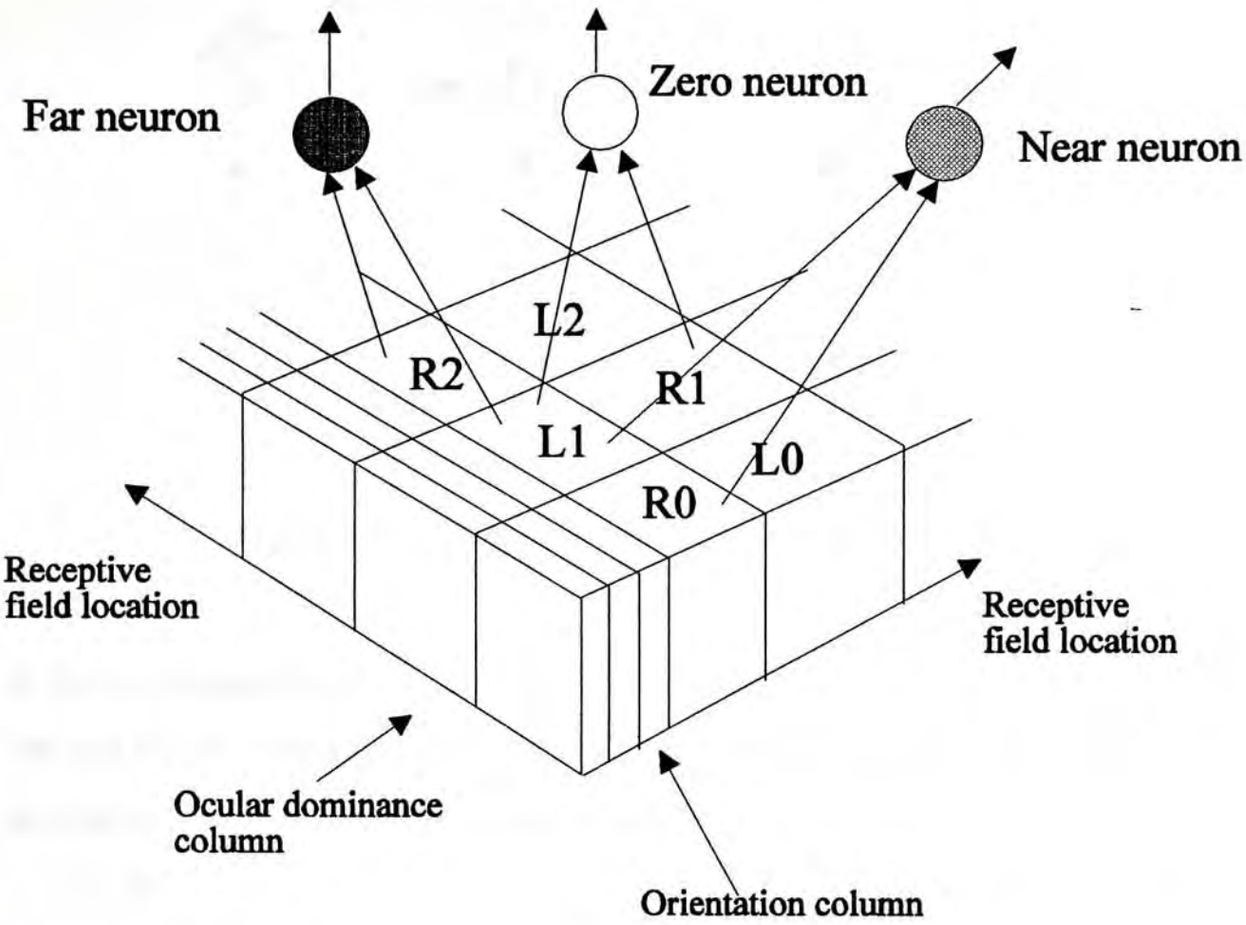


Figure 3.19 A diagrammatic picture of disaprity detection in PSVM.

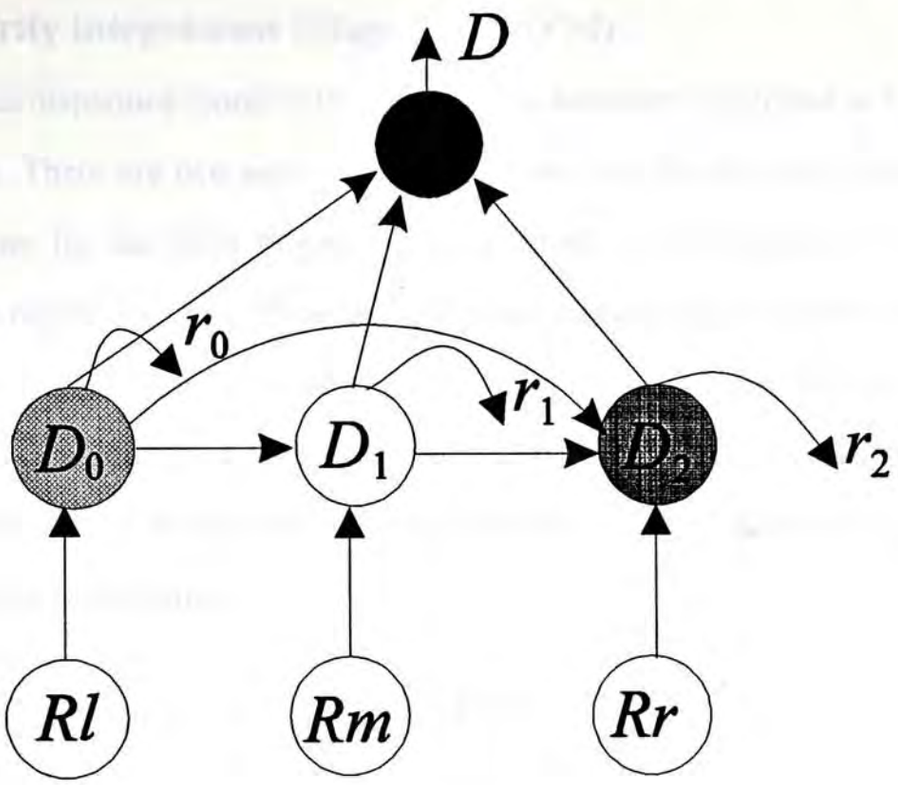


Figure 3.20 The wiring of the depth neurons in PSVM.

A further explanation of the connections of the neurons is shown in Figure 3.20. *Rl*, *Rm* and *Rr* are defined in section 3.4.7. The near neuron (D_2), the zero neuron (D_1) and the far neuron (D_0) are defined as follows:

$$D_0 = Rl \tag{3.24}$$

$$D_1 = \begin{cases} Rm, & if\ D_0 = 0 \\ 0, & otherwise \end{cases} \tag{3.25}$$

$$D_2 = \begin{cases} Rr, & if\ (D_0 = 0) \wedge (D_1 = 0) \\ 0, & otherwise \end{cases} \tag{3.26}$$

$$D = \begin{cases} D_0, & if\ D_0 \neq 0 \\ D_1, & if\ (D_0 = 0) \wedge (D_1 \neq 0) \\ -D_2, & if\ (D_0 = 0) \wedge (D_1 = 0) \end{cases} \tag{3.27}$$

$$r_t = \begin{cases} 1, & if\ D_t > 0 \\ 0, & otherwise \end{cases} \tag{3.28}$$

where $t = 0..2$.

3.5 Disparity integrations (Stage 3 of PSVM)

Local disparities found in the intermediate levels are integrated at the disparity integrations. There are two sets of local disparities, one for the long orientated lines and the other for the short orientation lines, which are the inputs of the disparity integrations model. The outputs of this model are unambiguous disparities. This model is shown in Figure 3.21. It consists of two processes, namely, the voter and the redistributor. The function of the voter is to select the most popular disparity (winner) within a specific area (voting area). The redistributor is used to redistribute the popular disparity to the voting area.

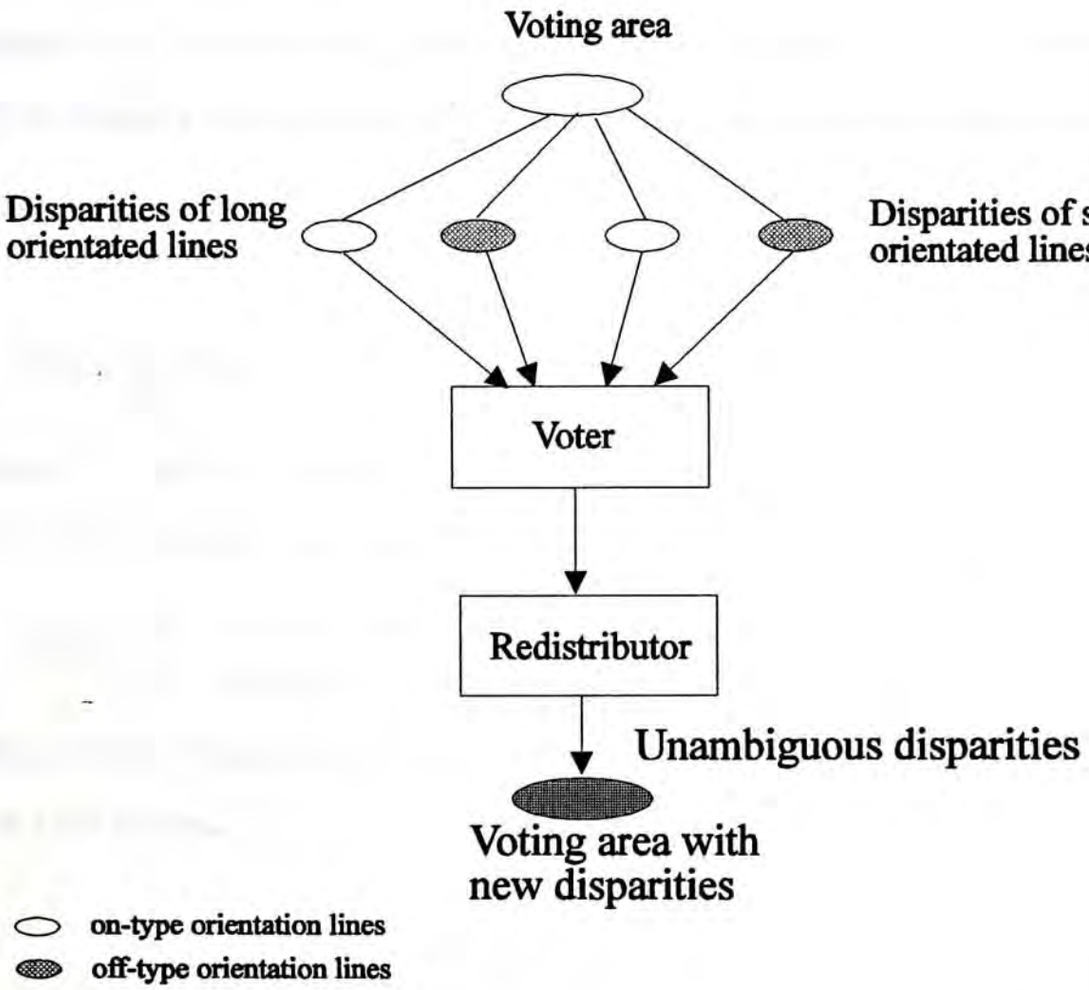


Figure 3.21 Block diagram of disparity integrations. The most popular disparity within a voting area will be selected by voter network (see section 3.5.1) and this popular disparity will be written back to that voting area by redistributor network (see section 3.5.2). Therefore, this area will have unambignous disparities.

3.5.1 The voter network

The voter network consists of two subnetworks, the P1 network and the P2 network. These subnetworks are identical except the threshold function of the network nodes. The structure of the voter network are shown in Figure. 3.22. Suppose that there are n neurons (voters) in the voting area, the nodes in the voter network can be defined as follows.

For $ND1$ neurons, the connections bewteen $ND1$ and D_t ($t = 0..2$) is given by

$$ND1_{ij} = \begin{cases} 1, & if\ (D_{t,i} = D_{t,j}) \wedge (i \neq j) \\ 0, & otherwise \end{cases} \tag{3.29}$$

where $0 \leq i, j \leq n$-1 and $D_{t,i}$ means the ith voter in the voting area. Note that D_t is one of the disparity neurons (near neuron, zero neruon and far neuron) that are defined in section 3.4.8.

For CD1 neurons,

$$CD1_i = \sum_{j=0}^{n} ND1_{ij} \tag{3.30}$$

where $i \neq j$ and $0 \leq i, j \leq n$-1.

For $ND2$ neurons, $0 \leq i, j \leq n$-1,

$$ND2_{ij} = \begin{cases} 1, & if\ (CD1_i \geq CD1_j) \wedge (i \neq j) \\ 0, & otherwise \end{cases} \tag{3.31}$$

where $CD1_i \neq 0$ and $CD1_j \neq 0$.

For CD2 neurons,

$$CD2_i = \begin{cases} 1, & if\ (\sum_{j=0}^{n-1} ND2_{ij}) = n-1 \\ 0, & otherwise \end{cases} \tag{3.32}$$

where $i \neq j$ and $0 \leq i, j \leq n$-1.

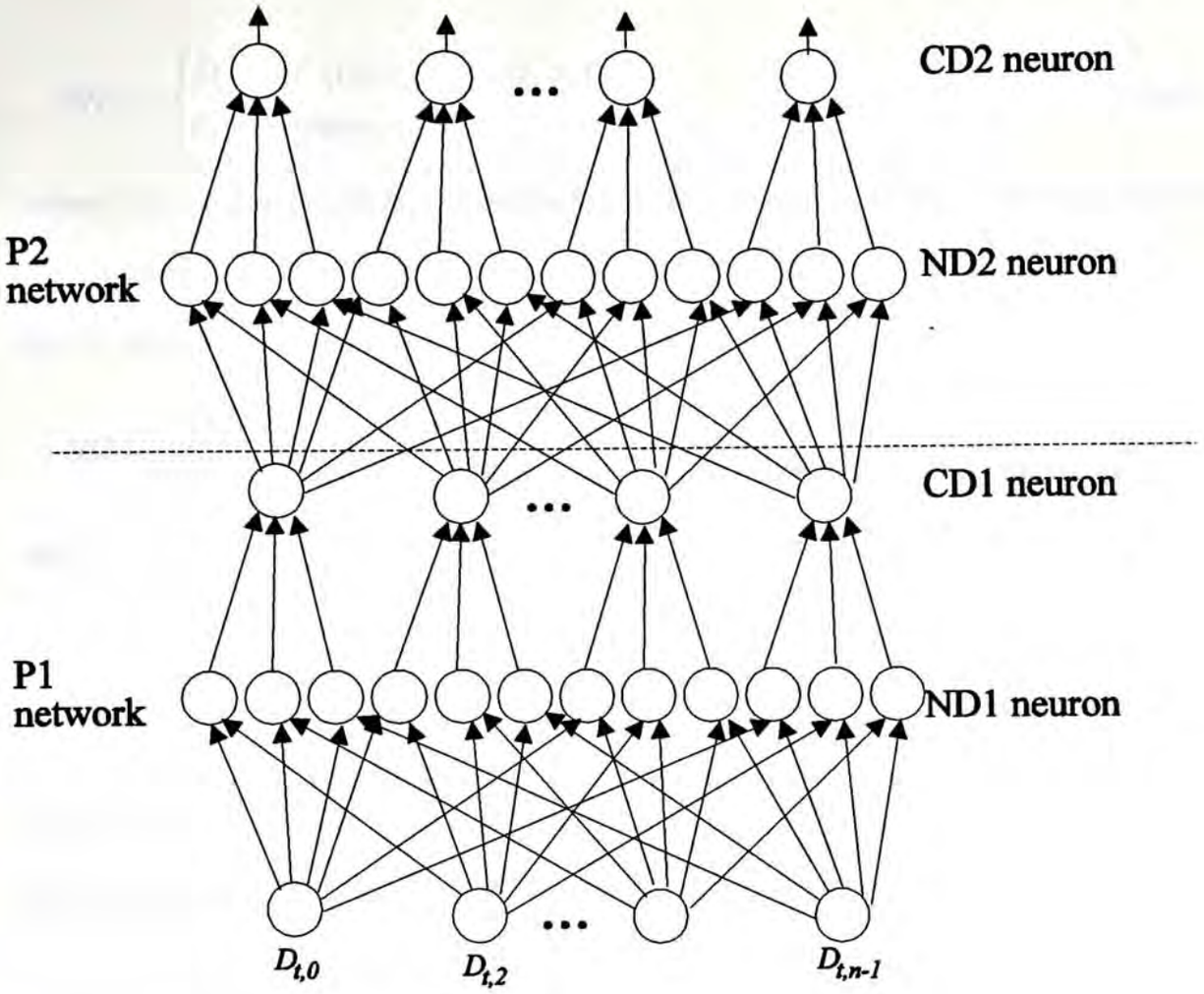


Figure 3.22 The voter network in disparity integrations

3.5.2 The redistributor network

The redistributor network selects the winner to renew the disparities in the voting area. The network consists of three parts, T1, P3 and T2. Figure 3.23 shows the connections of the network. P3 in the redistributor network is similar to P2 in the voter network (see section section 3.5.1). In T1, C is a control neuron that will let the redistributor network to select the larger disparity if no winner is selected in the voter network. The connections of T1 are

$$C = \begin{cases} 1, & if \ (\sum_{i=0}^{n-1} CD2_i) = 0 \\ 0, & otherwise \end{cases} \tag{3.33}$$

where n is the number of voter in the voting area, and

$$RN1_i = \begin{cases} D_{t,i}, & if\ (CD2_i = 1) \vee (C = 1) \\ 0, & otherwise \end{cases} \tag{3.34}$$

where $0 \le i, j \le n$-1, and $D_{t,i}$ is defined in section 3.4.8 and $CD2_i$ is the output of the voter network which is defined in section 3.5.1.

For P3 network, $0 \le i, j \le n$-1, the connections are

$$ND3_{ij} = \begin{cases} 1, & if\ (RN1_i \ge RN1_j) \wedge (i \ne j) \\ 0, & otherwise \end{cases} \tag{3.35}$$

and

$$CD3_i = \begin{cases} 1, & if\ (\sum_{j=0}^{n-1} ND3_{ij}) = n-1 \\ 0, & otherwise \end{cases} \tag{3.36}$$

where $i \ne j$.

The definitions of T2 network are

$$RN2_i = \begin{cases} D_{t,i}, & if\ CD3 = 1 \\ 0, & otherwise \end{cases} \tag{3.37}$$

$$NE_i = \begin{cases} RN2_i, & if\ (\sum_{j=0}^{n-1} NE_j) = 0 \wedge i \ne j \\ 0, & otherwise \end{cases} \tag{3.38}$$

$$DB = \sum_{i=0}^{n-1} NE_i \tag{3.39}$$

The old disparities in the voting area is replaced by DB.

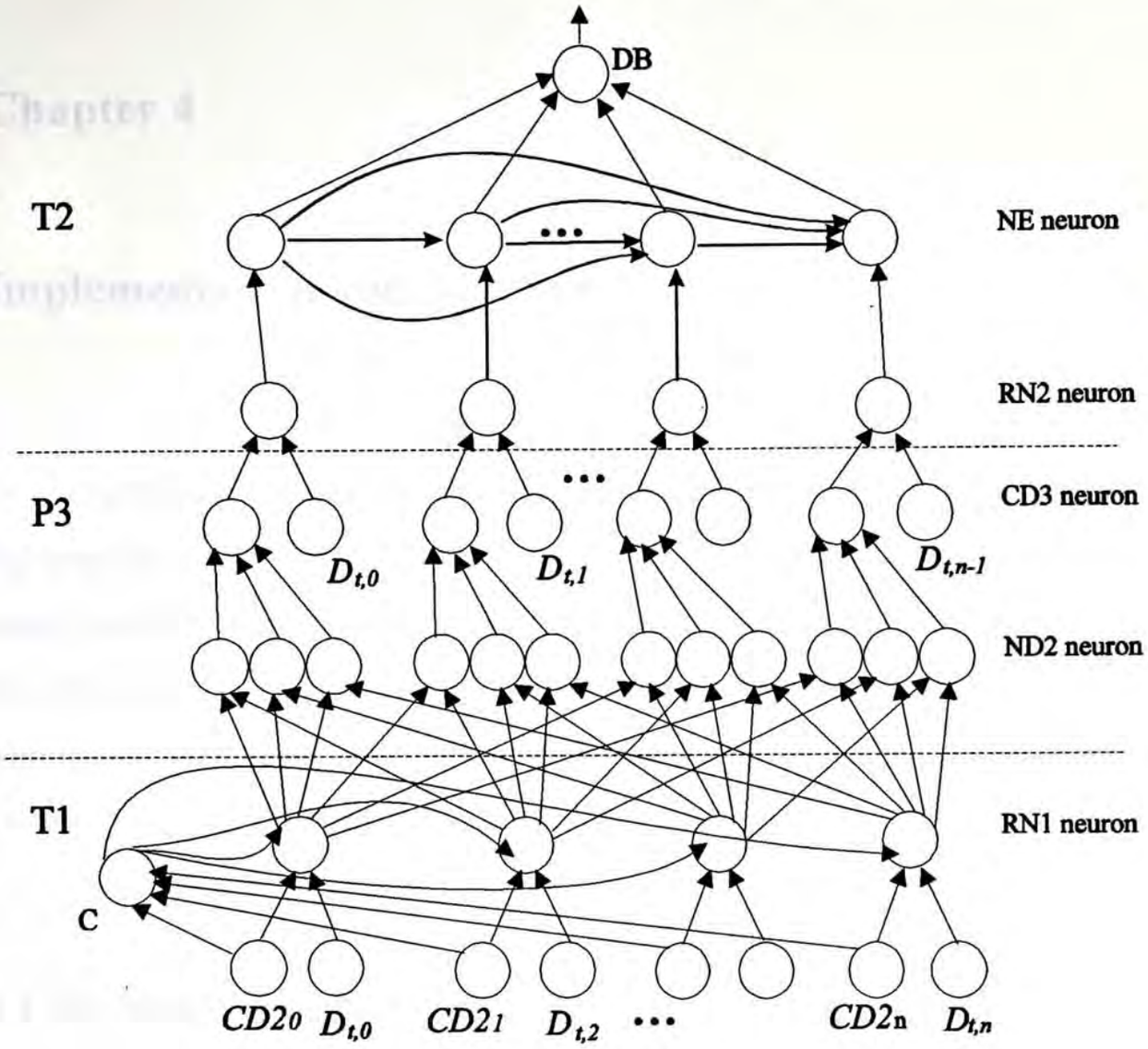


Figure 3.23 The redistributor network in disparity integrations.

3.6 Conculsion of chpater 3

In matching stereo images, there are two major problems: the extraction of image features and the determination of corresponding from a pair of images. PSVM extracts oriented lines from images as the matching features. To solve the second problem, PSVM uses a topographical mapping method, the hypercolumn structure, to achieve it. With the hypercolumn structure, the matching can be done in parallel and the representation of a three-dimensional object is easy and clear.

Chapter 4

Implementation and Analysis

In Chapter 3, a new stereo vision model called PSVM is proposed. PSVM addresses the stereo vision problem from the psychophysical view, it uses a special cortex structure called hypercolumn to handle the image information. A hypercolumn can serve as a processing module for a particular area of the retina [9, 18]. The binocular processing are localized at the hypercolumns and can be performed in parallel.

In this chapter, an detailed analysis of PSVM is discussed.

4.1 The imaging geometry of PSVM

For the class of stereo imaging geometry that we concern, the optical axes of the two cameras lie in the same plane, and all matching primitives appear on epipolar lines [1] (Figure 2.2). In a special case when the principal axes are parallel, all epipolar pairs will be lied horizontal and matching points will be found on corresponding rasters (Figure 2.3). This constraint restricts the search for possible matches to one dimension (horizontal dimension).

Like many other stereo vision models, PSVM uses this classical geometry to obtain epipolar constraint. However, it is quite difficult to get such ideal environment in practice. A pair of stereo images always have vertical misalignments. PSVM uses the fusional method similar to that of the human visual system to solved this misalignment problem. Instead of restricting the search for possible matches to one dimension, it matches possible primitives over an area (see Chapter 3).

4.2 Input

A pair of images, representing the right and left views of the scene, are inputted into PSVM. The input images has 256 gray-level. The primitives of the images are extracted by convolution. Although a DOG filter is suggested, it is possible to use other filters to extract edge information from the input images. Figure 4.1 shows two examples of output given by two different edge detectors. Figure 4.1(b) is the output of the Shaw's edge detector [33]. Figure 4.1(c) is the zero-crossings of the DOG image.

4.3 The hypercolumn construction

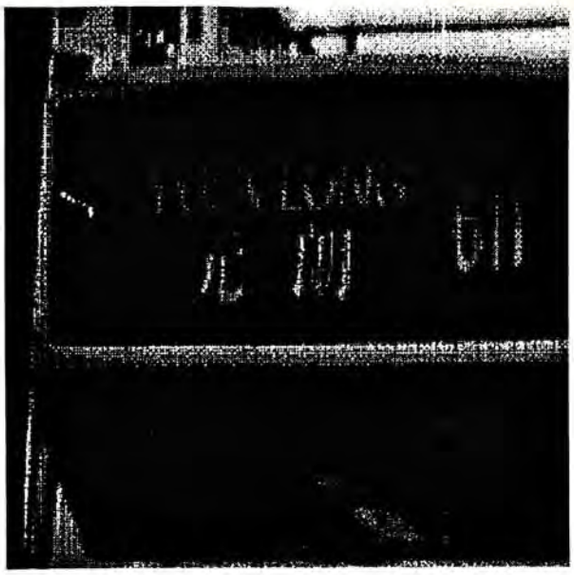
After edge detection, the oriented line segments are extracted and placed into the hypercolumns. As mention in Chapter 3, there are four classes of orientation columns: vertical , horizontal , left diagonal , and right diagonal. The constructions of different kinds of lines in different orientations are independent, and they can be implemented in parallel.

In PSVM, the hypercolumn is a cube structure (see Chapter 3). To account for this structure, in the present implementation, the hypercolumn is a three dimensional array structure and PSVM uses a pair of these arrays to represent the left and right ocular dominance hypercolumns. Different portions of these arrays represent different features in different locations of the retina image.

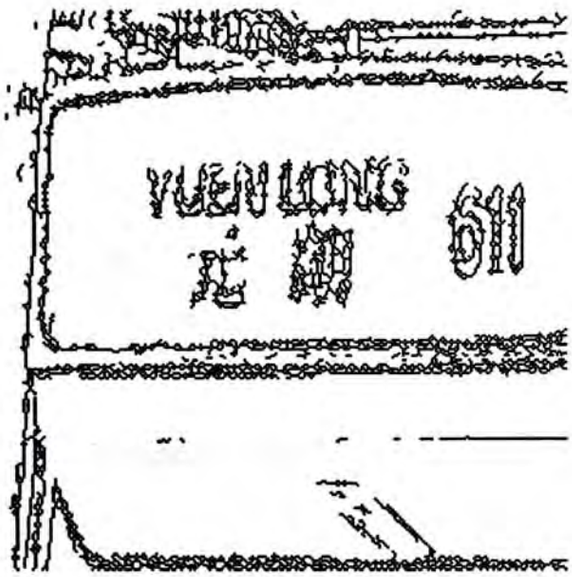
4.4 Analysis of matching mechanism in PSVM

The correspondence problem is solved by the local disparity detection module in PSVM. The detection of local disparity comprises three steps:

1. Oriented line constructions.
2. Oriented lines matching.
3. Disparities classification (far, near and zero disparity).



(a) Sample image



(b) The output of Shaw's edge detector.

(c) The zero-crossings of the DOG image.

Figure 4.1 The results of the Shaw's edge detector and the DOG operator.

They are the main steps of fusional method in the human visual system which solves the correspondence problem. A pair of left and right oriented lines are matched if their orientations are same or roughly same and their length are equal or nearly equal. Further more, their disparities must be within the fusional area (Panum's fusional area). Note that there are two main parameters for fusion: orientation differences (*da*) and length differences (*dl*) of the lines. If a pair of potential candidates meets these two criteria, the left and the right can be fused.

4.4.1 Fusional condition

As mentioned above, there are two criteria, d_l and d_a for two potential candidates to fuse. They can be defined as

$$dl = \frac{|d_l - d_r|}{d_l + d_r} \tag{1}$$

where d_l is the length of the left image line, and d_r is the length of the right image line, and

$$da = |a_l - a_r| \tag{2}$$

where a_l is the orientation of the left image line and a_r is the orientation of the right image line. With this two criteria, the fusional condition is set up as follows

$$f = \begin{cases} 1; & if\ fp < 1 \\ 0; & otherwise \end{cases} \tag{3}$$

and fp is given by

$$fp = \left| \frac{dl}{DR} - \frac{da}{AR} \right| \tag{4}$$

where DR and AR are, respectively, the threshold values of the difference of angle and the difference of length between two oriented lines. However, it is not enough if a fired disparity cell is only controlled by the fusional condition.

4.4.2 Disparity detection

After the potential candidates are fused, PSVM selects the right disparity and let the corresponding disparity cell fire. As mentioned in Chapter 3, the fusional area in PSVM is a rectangle area and contains eight elements. There are cases that a feature on the left retina image may fuse more than one features on the right image. Figure 4.2 shows an example of such case. A competition model has been used to select the correct disparity,. This is described in Figure 4.3. Note that only four candidates are listed in Fingure 4.3. Within a pool (near disparity, far disparity or zero disparity), the candidate will win if its fusional value (fp) is a minimum. It can be formulate as

$$w = \min(fp_i) \tag{5}$$

where $i \in F$. F is the set that contains all possible candidates and is given by

$$F = \{x | f = 1\} \tag{6}$$

where f is defined in section 4.1. After the competition, there are only three possible depth cells (near disparity cell, far disparity cell and zero disparity cell) to be considered (see Chapter 3). PSVM uses a stronger-suppress-weaker model to select the depth cell. The connections of this model are described in Chapter 3.

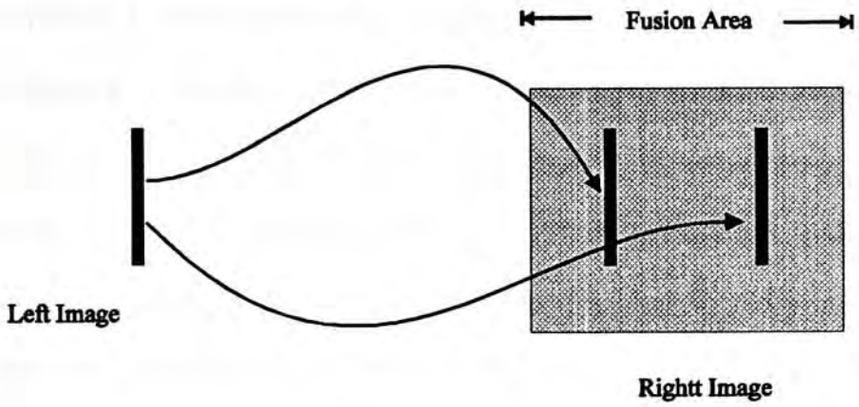


Figure 4.2 Example of multi-fusion.

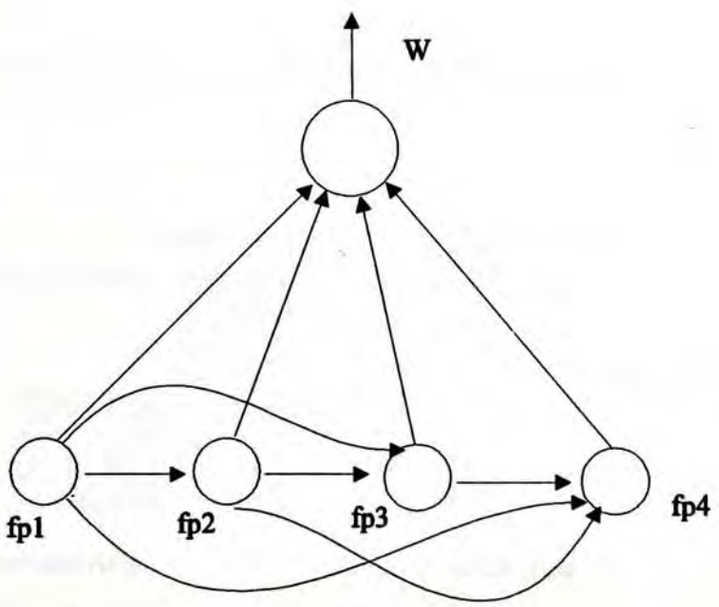


Figure 4.3 A competition model to solve the ambiguous disparities.

4.5. Matching rules in PSVM

Similar to the fusional condition discussed in the earlier section, PSVM also implies other matching constraints. In this section, the relation between the constraints are discussed.

4.5.1 The ordering constraint

As described in Chapter 2, stereo projection at most of the time preserves the order of edges extracted from the two images. Therefore, matches which maintain the ordering are preferred. This constrain is implied in PSVM at the first and the finial selection of a disparity. The first selection of disparity is to select a disparity in a pool of far, near or zero disparity. Figure 4.4 describes how these three pools correspond to the hypercolumn structure. The fusional area is divided into the left fusional area, the middle fusional area and the right fusional area (see Chapter 3). PSVM uses different functions to pick out the disparity in different fusional areas. In the left fusional area, PSVM chooses the cell that has the maximum disparity

$$D_l = \max(X_l - X_r) \tag{7}$$

While in the right fusional area, the cell that has minimum disparity is selected

$$D_r = \min(X_l - X_r) \tag{8}$$

As for the middle fusional area, the displacement will be selected

$$D_m = X_l - X_r \tag{9}$$

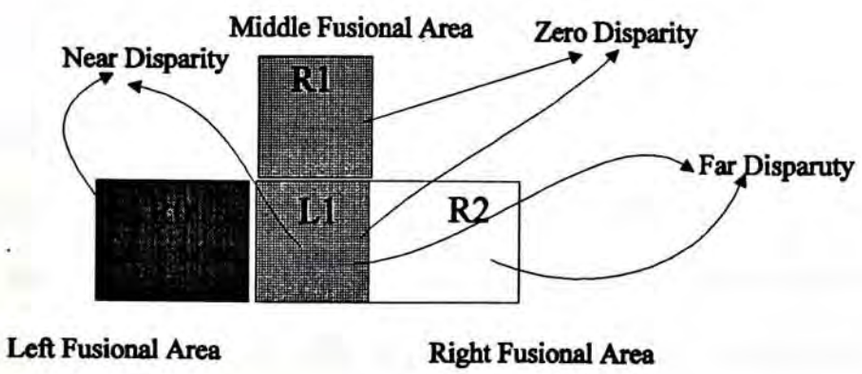


Figure 4.4 The detections of near disparity, zero disaprity and far disparity in PSVM.

4.5.2 The uniqueness constraint

It is true that each image primitive only matches one corresponding primitive in another image. In PSVM, this constrain is implemented by a feed back control loop. Figure 4.5 shows a complete version of the matching mechanism in PSVM. The signal *r0*, *r1* and *r2* are the control signals. They prevent the chosen cells from double matching. For example, if H2 is the winner in the disparity competition, *r0* will excite C2, and hence, H2 will be forbidden in the next matching period.

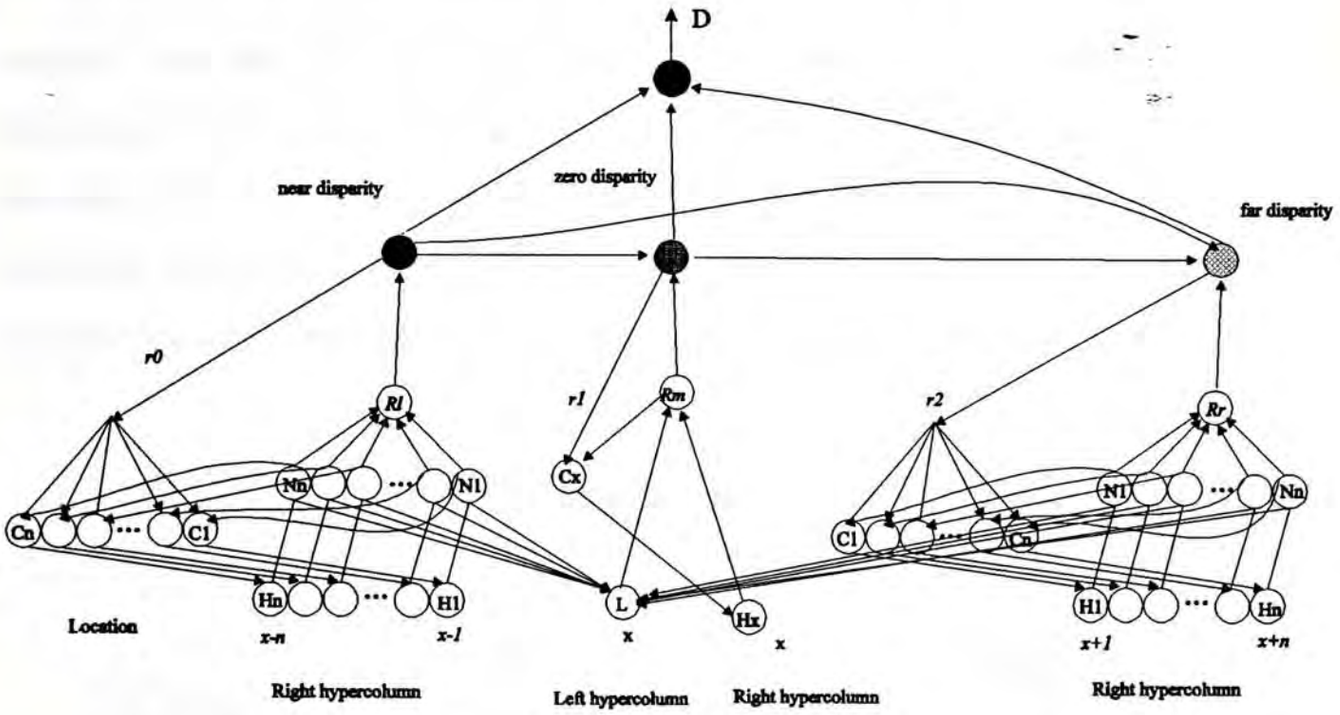


Figure 4.5 Matching mechanism in PSVM.

4.5.3 The figural continuity constraint

The figural continuity constraint is suggested by Mayhew and Frisby [25]. They suggest that smoothness of disparity is only required along contours (see Chapter 2 for details). This constraint is naturally implied in PSVM. It is because PSVM uses oriented lines as matching features. The oriented line is a contour by itself. Furthermore, the construction of long oriented lines also uses this constraint for the disparities along a oriented line is smoothness.

4.5.4 The smoothness assumption

The smoothness assumption is based on the observation that surfaces are smooth except at the discontinuities in depth. This assumption is implemented by disparity integrations. When disparities are integrated, PSVM always assumes that disparities are smooth within a specific area where the long oriented lines and short oriented lines join. In Chapter 3, this area is called the voting area. In the voting area, PSVM finds the most proper disparity. Once the popular disparity is found, it will be assigned over this area. This idea is illustrated by Figure 4.6. In Chapter 3, two networks, the voter network and the redistributor network, are proposed to implement this idea. The voter network is used to find the popular disparity. If there are no popular disparity, the disparity that has greater value will be selected. The redistributor network is used to renew the disparity in the voting area with popular value.

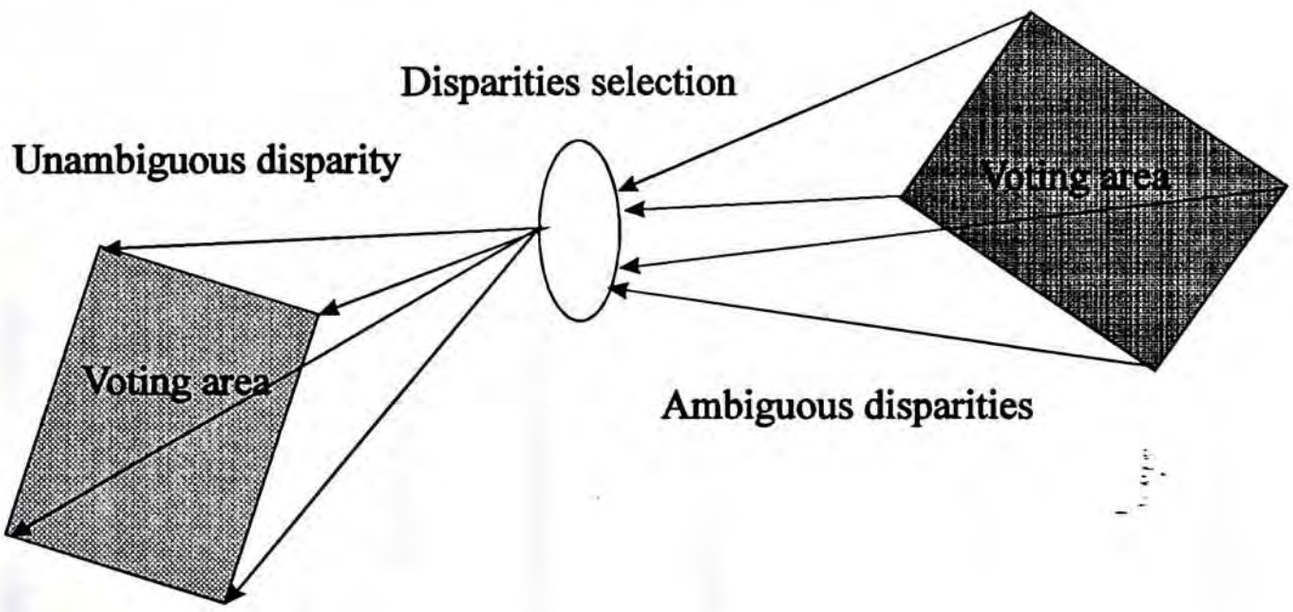


Figure 4.6 The model of the unambiguous disparity selection.

4.6. Use multi-lengths of oriented line to solve the occlusion problem

Different from other stereo vision models, PSVM uses oriented line with multiple lengths for matching. The long oriented lines give rough information of an object while the short oriented lines give fine details.

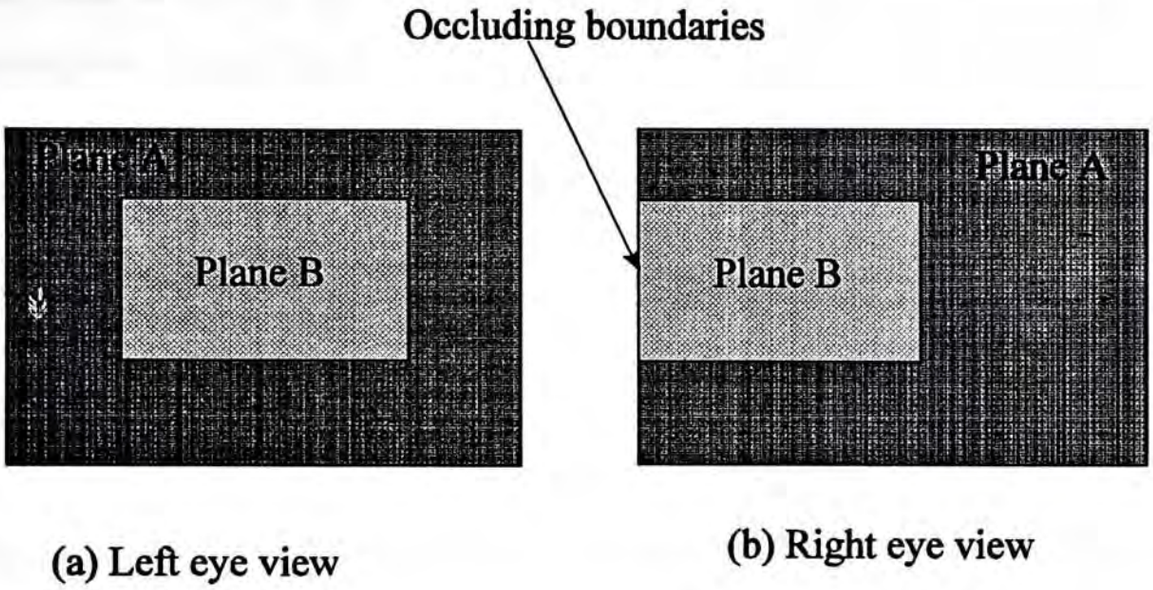


Figure 4.7 An occluding boundaries example.

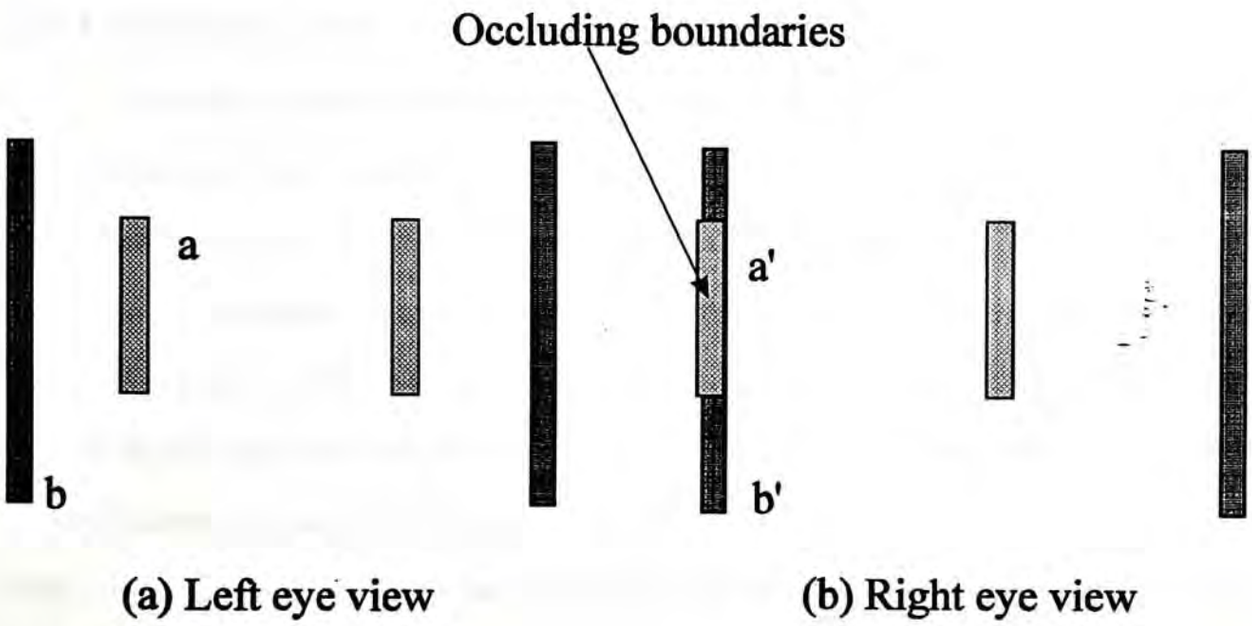


Figure 4.8 The occluding boundaries' matching in PSVM.

Matching errors always occur along discontinuities in depth, especially at occluding boundaries. An occluding boundaries example is shown in Figure 4.7. Suppose that there are two planes A and B and the object is viewed from the right top corner. The left eye view and right eye view are shown in Figure 4.7(a) and 4.7(b). A sharp break in disparity between two planes A and B is occurred. Now, let's consider vertical lines matching (Figure 4.8). In the long lines matching, vertical line (b) in left view (Figure 4.8(a)) matches vertical line (b') and (a) is missed. However, in short lines matching, vertical line (a) matches (a') and (b) is missed. When these two disparity maps are integrated, a full disparity description of the object is obtained.

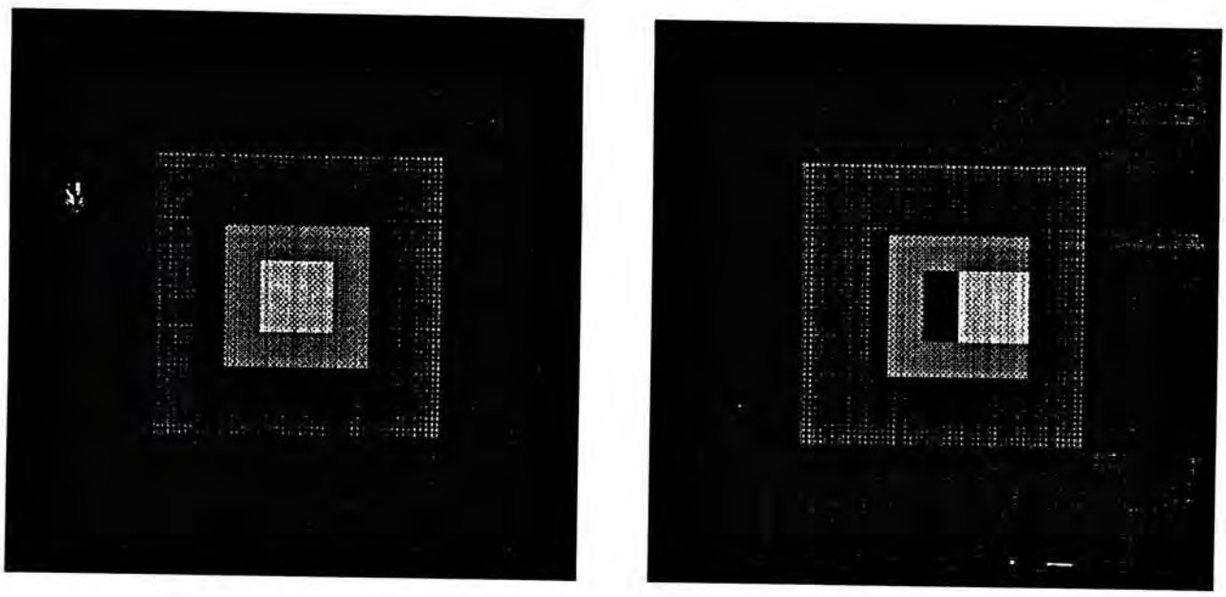
4.7 Performance of PSVM

In this section, two different types of stereo images, artificial images and natural images, are examined so that we can evaluate the performance of PSVM and its potential applicability to automated stereo acquisition of depth information in robotics and cartography.

4.7.1 Artificial scene

Since the disparity values of artificial images are known, it is easier to assess the correctness of the model's performance. The first example of an artificial scene is shown in Figure 4.9. The scene is composed of three square planes. The first layer is the larger square plane which is on the bottom and the third layer is a smaller one that is put on the top. In this example, all the planes are shifted to the right while they are viewed by the right eye. Note that occlusion is occurred in this case. The right vertical edge of the top planar is overlapped with the right vertical edge of the second planar. The edge detection of the stereo images are shown in Figure 4.10(a) and Figure 4.10(b). In this example, Shaw's edge detector [33] is used to extract edge information as a demonstration of the possibility of using another edge operators in PSVM. After edge detection, PSVM divides the edges into two different types, the on-type edges and the off-type edges (see Chapter 2). Figure 4.10(c) and Figure 4.10(d) show the left

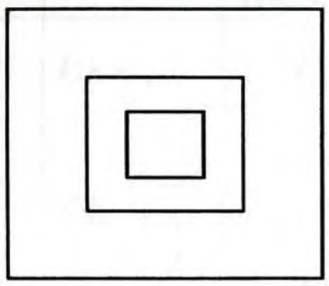
and right on-type edges and the left off-type edges and the right off-type edge are shown in Figure 4.10(e) and Figure 4.10(f).



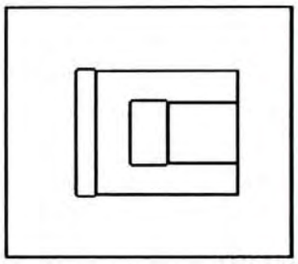
(a) Left image (b) Right image

Figure 4.9 A pair of artifical stereo images.

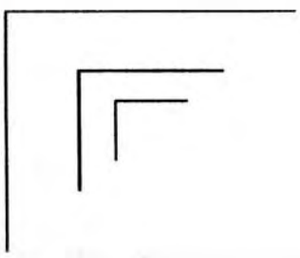
As mentioned earlier, PSVM uses lines of different lengths in matching to obtain full disparity information of an object. Lines of different lengths are extracted after the types of edges are detected. The total of short edges of on-type and off-type are 224 and 224 while the long edges of on-type and off-type are 6 and 6. In this example, the length of a short edge is less or equal to 3 pixels and a long edge is greater than 3 pixels. It should be noted that edges in PSVM are divided into the horizontal, the vertical, the left diagonal and the right diagonal lines (see Chapter 3). Table 1 lists the statistics of various oriented lines. The left diagonal and the right diagonal lines are not listed because they are not existed in this artificial scene.



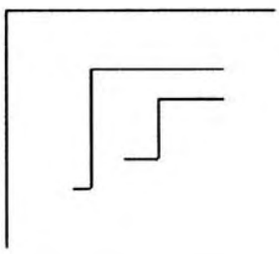
(a) The edge detection of the left image



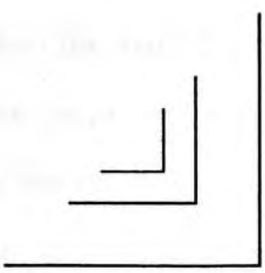
(b) The edge detection of the right image.



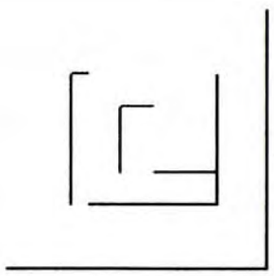
(a) On-type edges of the left image



(b) On-type edges of the right image



(c) Off-type edges of the left image.



(c) Off-type edges of the right image.

Figure 4.10 The edges of the artifical scene.

		Left View			Right View		
		Horizontal	Vertical	Total	Horizontal	Vertical	Total
Short Oriented Line (≤ 3 pixels)	On-type	112	112	448	124	112	504
	Off-type	112	112		124	144	
Long Oriented Lines (>3 pixels)	On-type	3	3	12	5	3	17
	Off-type	3	3		5	4	

Table 1

After the oriented lines are extracted, the matching process starts. The matching statistics of miscellaneous oriented lines are listed in Table 2. The matched short oriented lines are 432 and the unmatched ones are 16. The total number of matched short oriented lines is 96.4%. On the other hand, the long oriented lines have 11 matched and 1 unmatched result. The total matching percentage is 91.7%. There are unmatched oriented lines because of the occluding boundaries (see the description in section 6). When the matching process is finished, a disparity map is obtained. The different disparity maps produced after matching are shown in Figure 4.11. The disparity maps of the short oriented lines is shown in Figure 4.11(a). The result of the long oriented lines is shown in Figure 4.11(b). Note that the disparity maps obtained by PSVM are sparse arrays. It is because PSVM detects only the disparities of edges. From Figure 4.11, it can see that some edges in the figure is lost. To obtain a full disparity map of the object, PSVM merges all these disparity information into one. This is done by the disparity integrations of PSVM (see Chapter 3). The result of the combined disparity map is shown in Figure 4.12(a). There is no information lost and

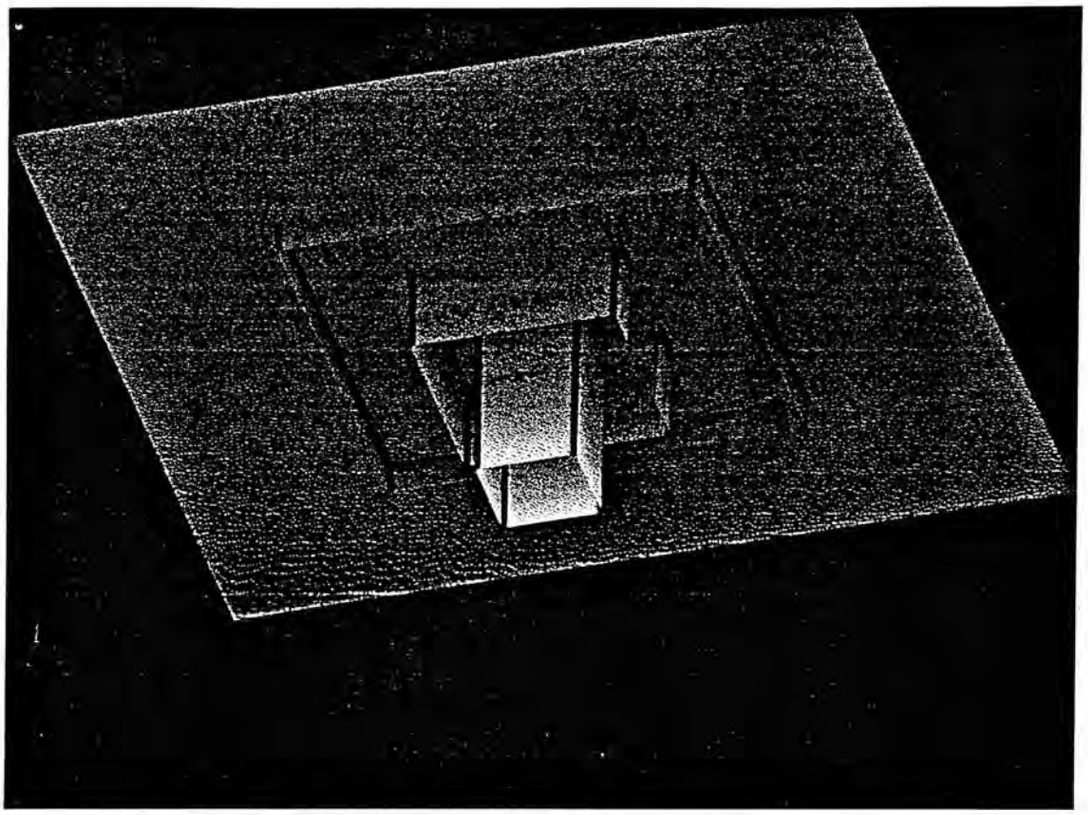
the occluding boundaries are recovered. Once the disparity map of the scene has been obtained, it can be used to reconstructe the scene (disparity interpretation). The interpolation of this disparity is shown in Figure 4.12(b). The disparity interpretation will not be discussed in this thesis because it is our another research topic.

			Matched	Unmatched	Matched Percentage	Total Matched Percentage
Short Oriented Line (≤ 3 pixels)	On-type	Horizontal	112	0	100%	96.4%
		Vertical	112	0	100%	
	Off-type	Horizontal	112	0	100%	
		Vertical	96	16	85.7%	
Long Oriented Line (> 3 pixels)	On-type	Horizontal	3	0	100%	91.7%
		Vertical	3	0	100%	
	Off-type	Horizontal	3	0	100%	
		Vertical	2	1	66.7%	

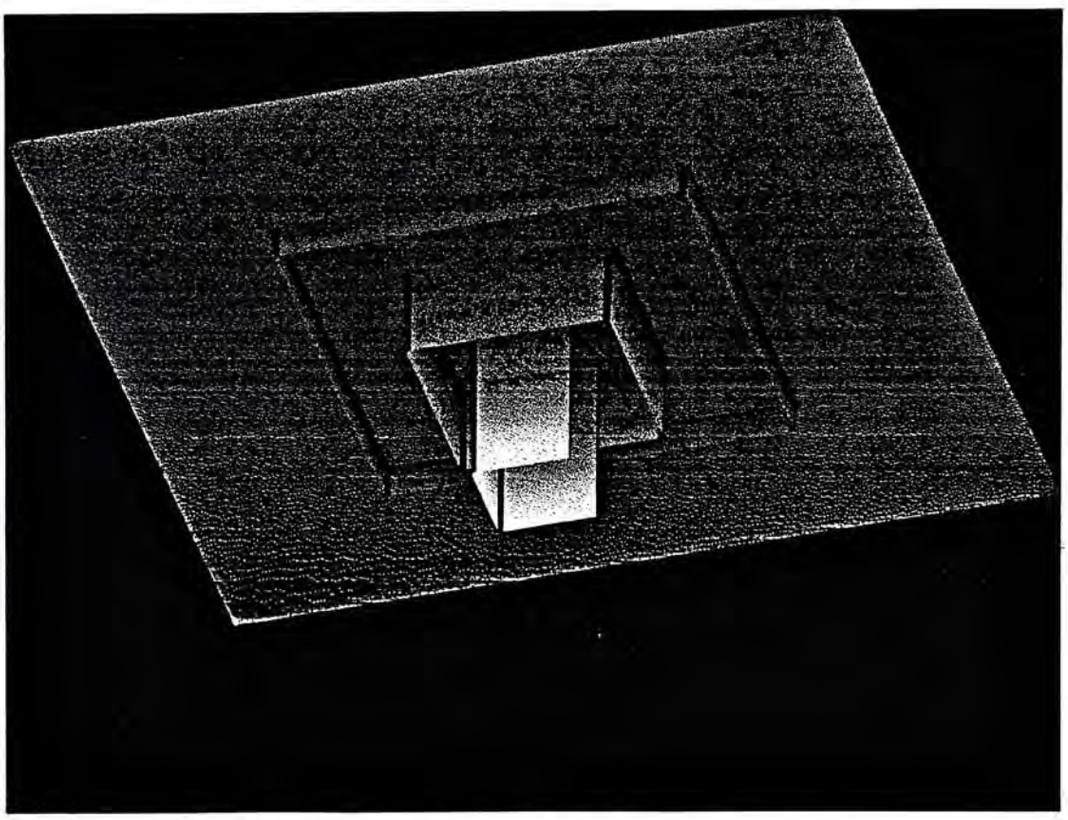
Table 2

4.7.2 Natural images

The first example of a pair of stereo images is shown in Figure 4.13. It is a scene of a laboratory. The scene described a box placed on a mouse pad. The size of the images is 300 by 300 pixels. The CCD camera with a 512 by 512 pixels and 256 gray-level is used to take these images. The camera took the shot from the top view of the objects.

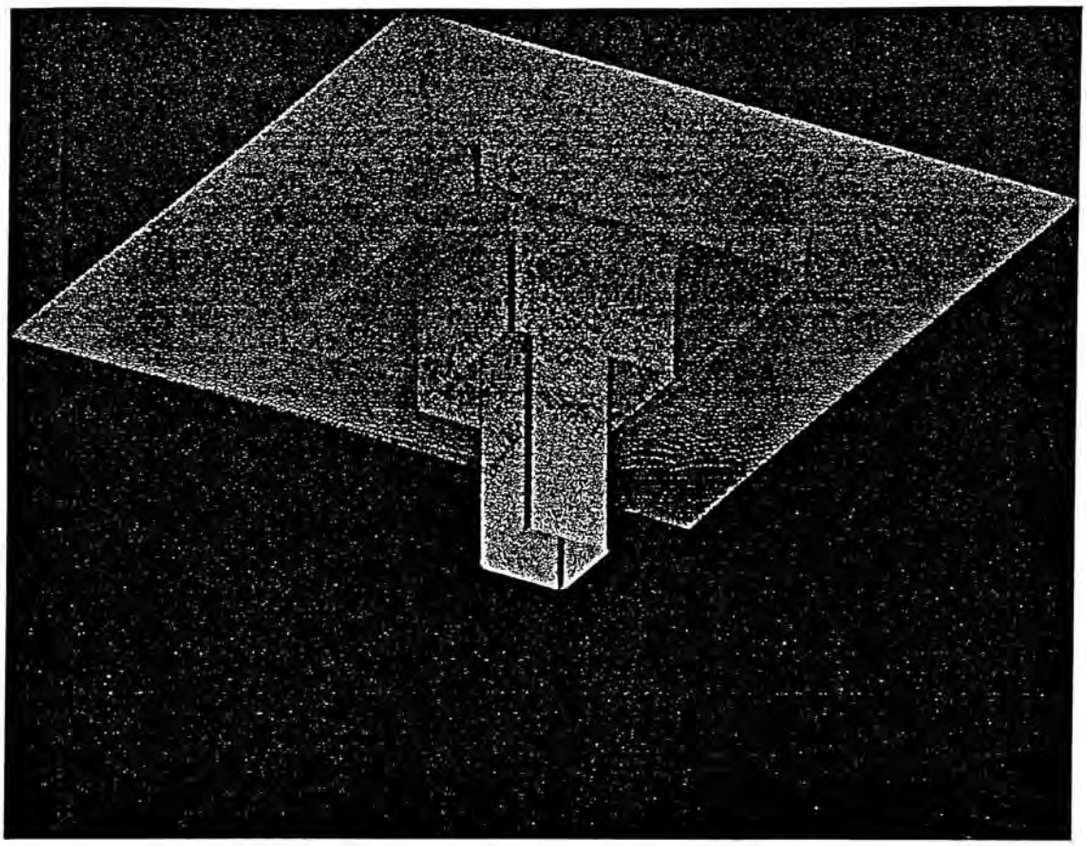


(a) The disparity map of the short oriented lines .



(b) The disparity map of the long oriented lines.

Figure 4.11 The disparity map of the short and long oriented lines of the artifical scene.



(a) The disparity map of the artifical scene.

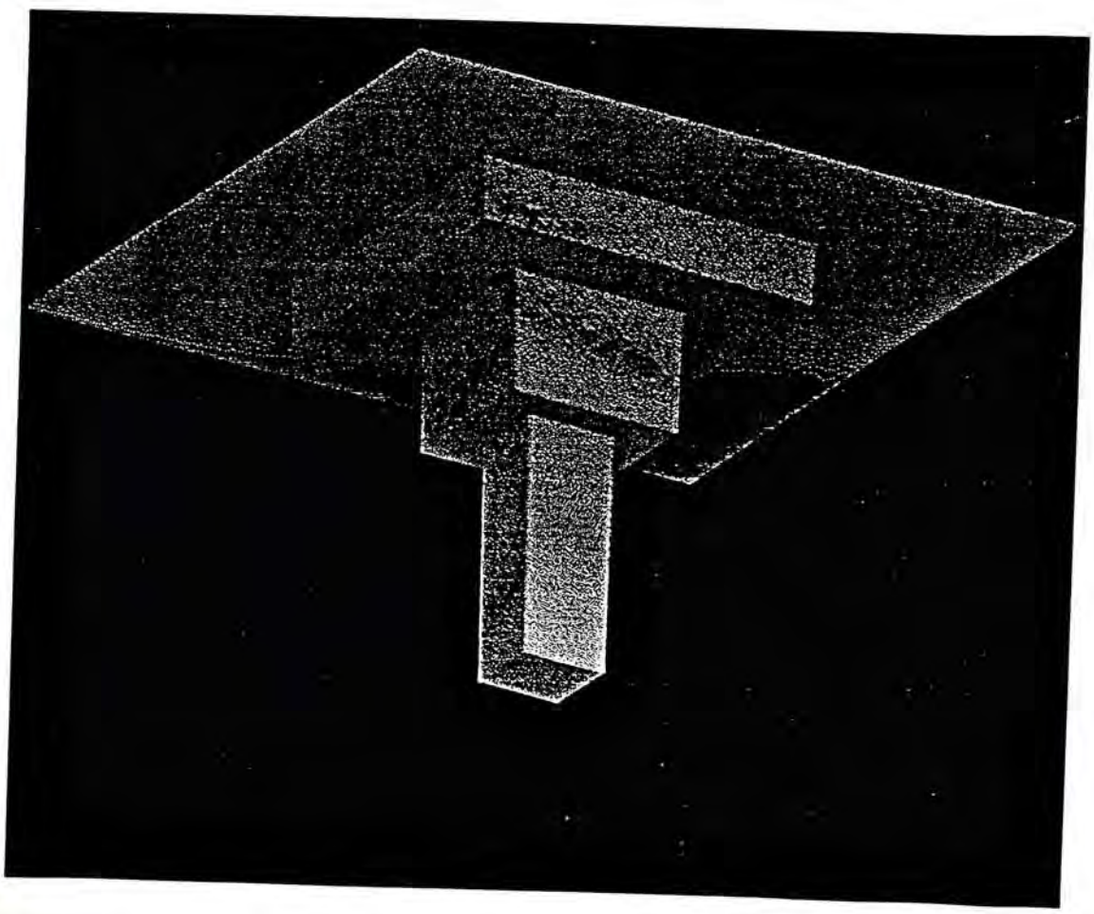
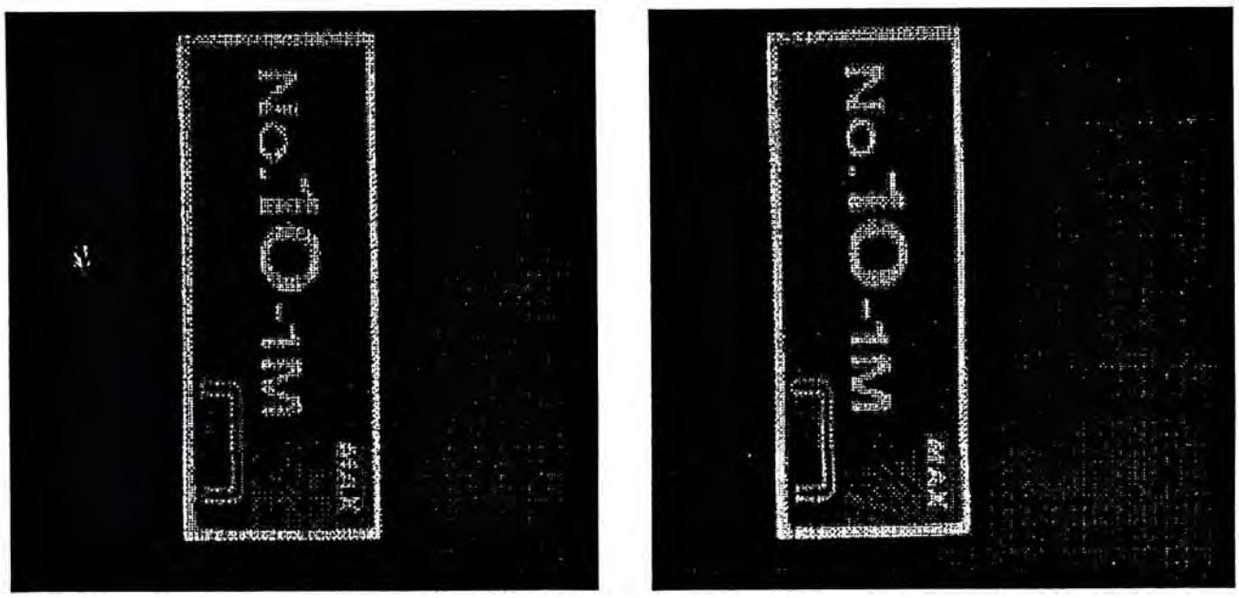


Figure 4.12 The disparity map and interpolation of the artifical scene.

The left and right images were convoluted with a DOG filter. The first Gaussian operator has a standard deviation of 1.5 pixels, and the second one has 3.6 pixels. Size of the actual mask is 31 by 31 pixels. After the DOG convolutions, the zero-crossing detection is applied. The results are shown in Figure 4.14. From this figure, it is noted that the zero-crossing is sensitive to noise. This is another issue in computer vision. In fact, PSVM does not limit the edge detection to DOG or $\nabla^2 G$. It can use other edge detection techniques, such as the Shaw's edge detector [33] or the Canny edge detector [4]. A demonstration of using the Shaw's edge detector has been shown in earlier paragraph at this section.

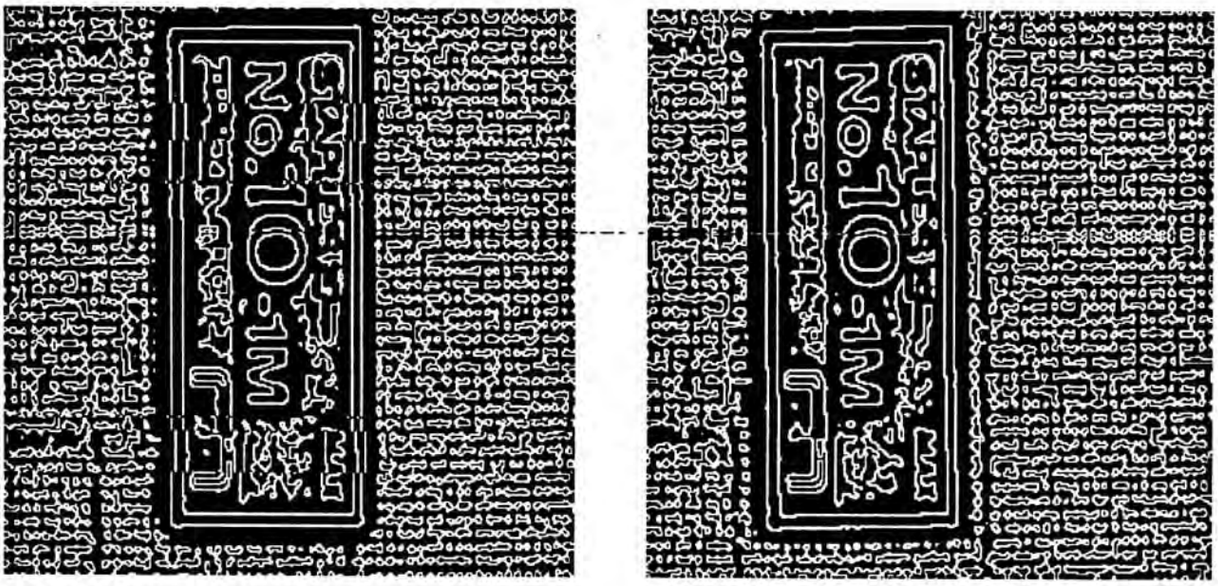


(a) Left image (b) Right image

Figure 4.13 The stereo images are of a scene of a box that was placed on a mouse pad.

Although the cameras are carefully mounted, it is very difficult to obtain a pair of stereo images that have perfected alignment. The vertical disparity in this set of images is ±2 lines. As mentioned in Chapter 3 (also see section 1 in this chapter), PSVM uses a fusional method to achieve the matching. Therefore, these misalignments

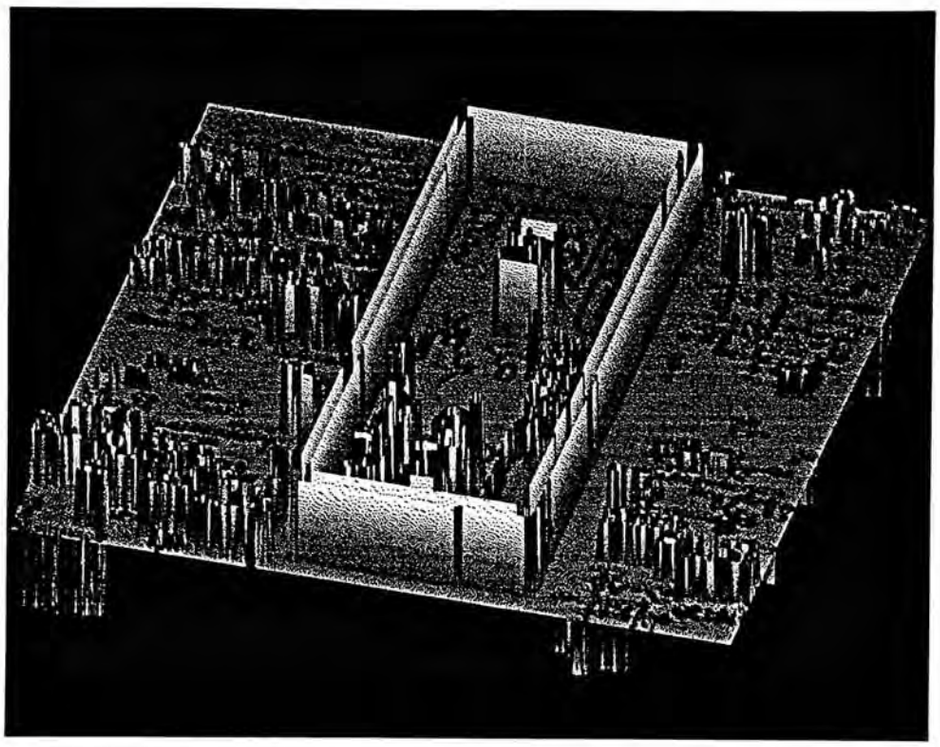
can be handled by PSVM. Moreover, PSVM groups edges into two types, the on-type edges and the off type edges, and several orientations, including the horizontal orientation, the vertical orientation, the left diagonal orientation and the right diagonal orientation. One of the advantages of this grouping is that the matching candidates are reduced, and hence the correctness of the matchings are increased.



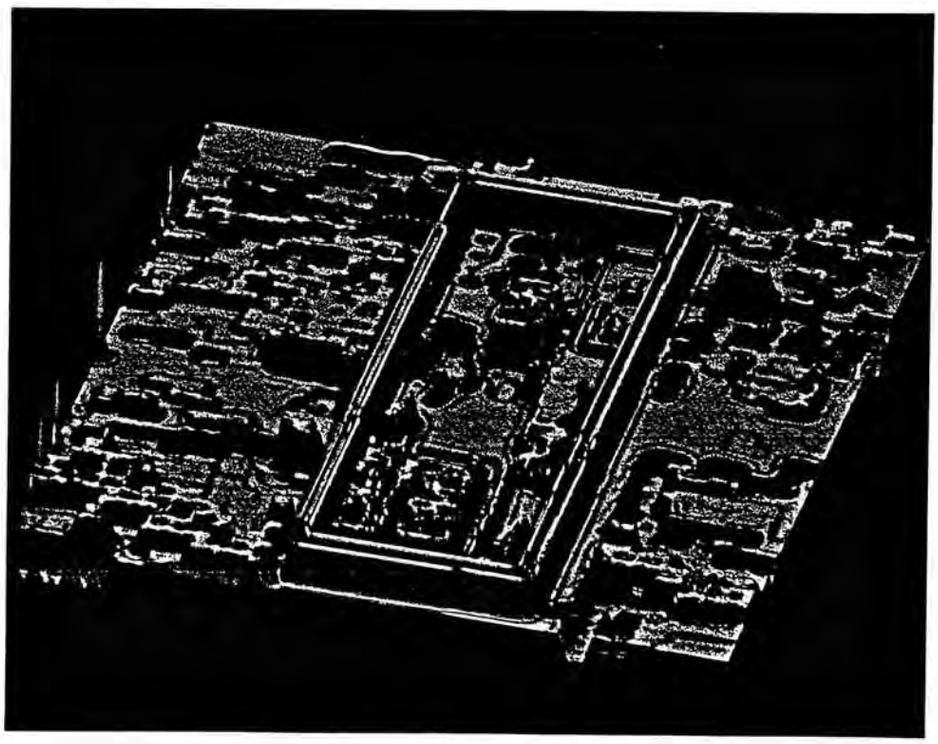
(a) The edge detection of the left image. (b) The edge detection of the right image.

Figure 4.14 The edge detection of the scene.

The result of merging the disparity maps into a single disparity map is shown in Figure 4.15(a). Some abrupt changes are observed in above figures. These irregular changes are due to the noise as noted earlier. However, these noise disparity information are not attached to the main object in PSVM. It can be seen that the contours of the box are correct although there are many noises in the input images. The interpolation of this disparity map is shown in Figure 4.15(b).



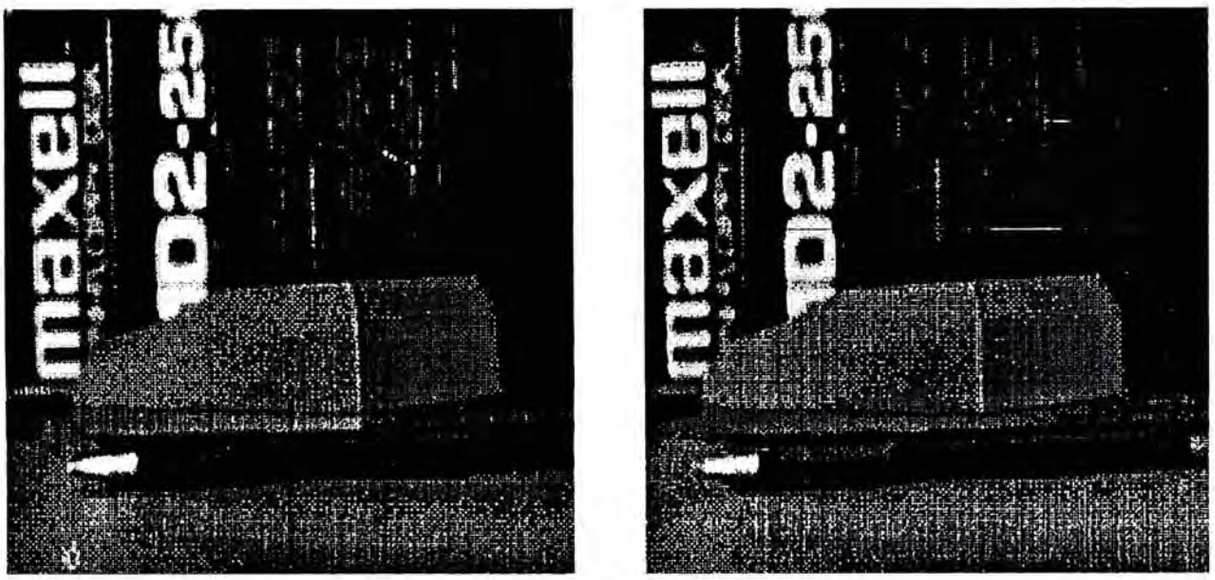
(a) The disparity map of the first pair of stereo images.



(b) The interpolation of the disparity map.

Figure 4.15 The disparity map and the recovered disparityies of the scene.

The second example of a laboratory scene is shown in Figure 4.16. The scene is composed of a disk box, a mouse and a pen. The pen is placed in front of the mouse and the disk box is put behind the mouse. The images were taken with the same CCD camera that used in the first example. The size of the images is 256 by 256 pixels with 256 gray-level. The cameras, in this example, were placed in front of the pen. The vertical disparity in this set of images is in the range of ±3 lines.



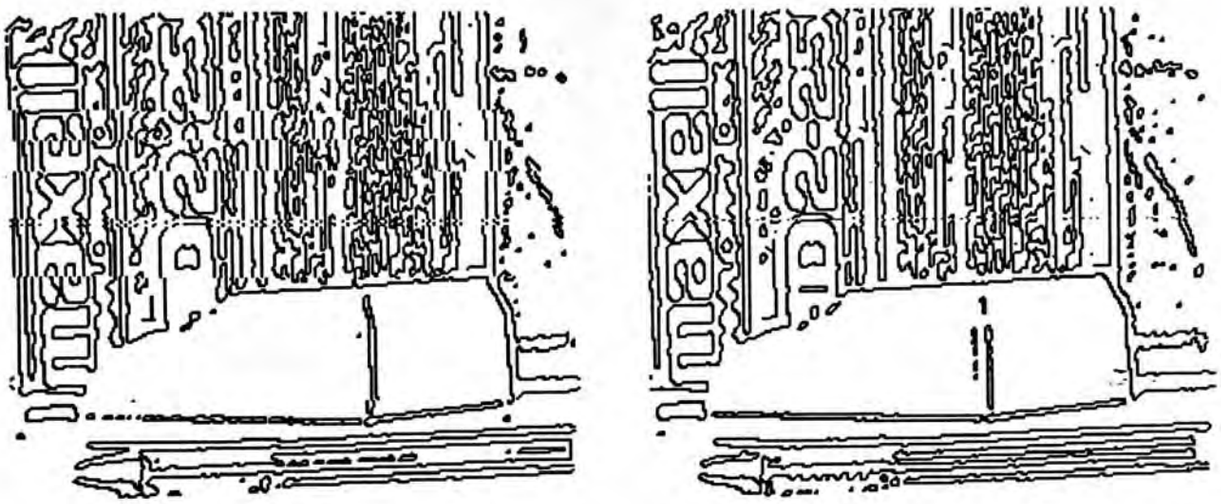
(a) The left image of the scene. (b) The right image of the scene.

Figure 4.16 The stero images are of a scene of a mouse.

A $\nabla^2 G$ filter (σ = 1.5 pixels) is used to convolute this scene. The zero-crossing results are shown in Figure 4.17. Once more, much noise are detected. The noise have not been removed and are used as a test for the model.

The final disparity map and the interpolation of the scene is shown in Figure 4.18. The scene is rotated to let the display look clearer. The pen and the mouse are placed at the upper image in Figure 4.18. From the disparity map, it can be seen qualitatively that the pen is much higher than the mouse. This means that the disparity values of the pen are larger than the mouse, and it is closer to the viewer. From the same map, it can be seen that the disparity values of the mouse are also larger then the disk box. It is reasonable that the disparity values of the disk box are not zero since

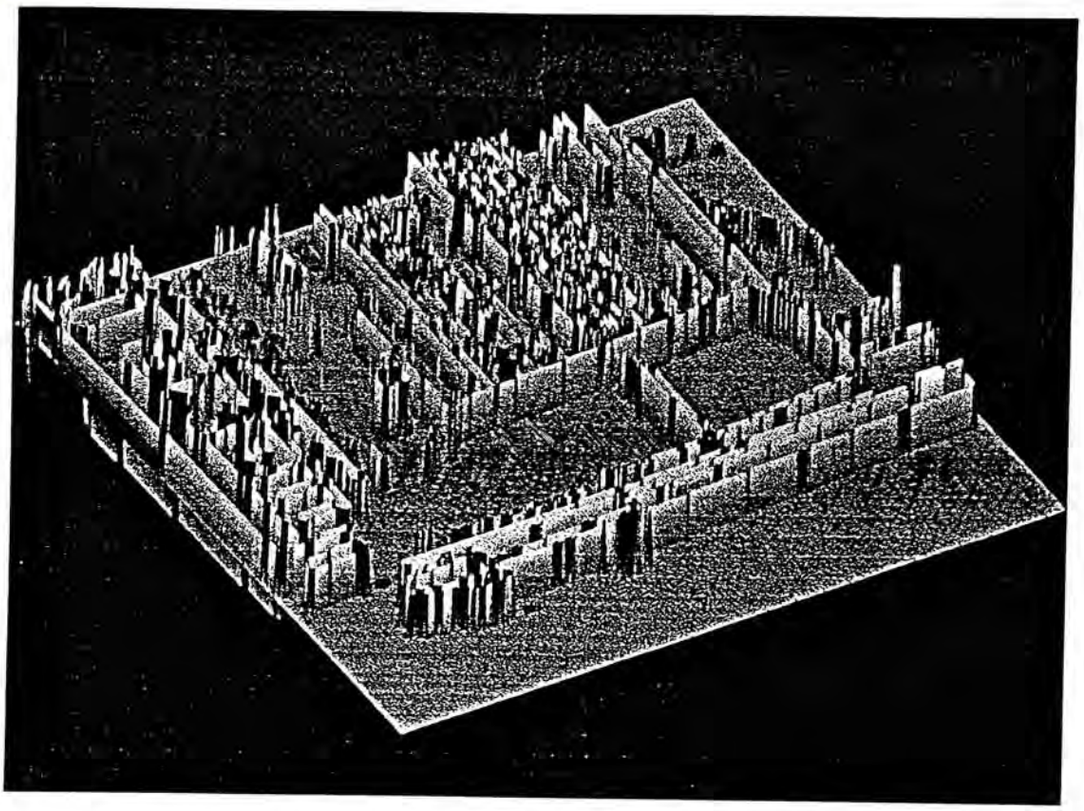
the imaging geometry of this set of images is a parallel epipolar geometry (see Chapter 2), the disk box is not far enough to create the zero disparity values.



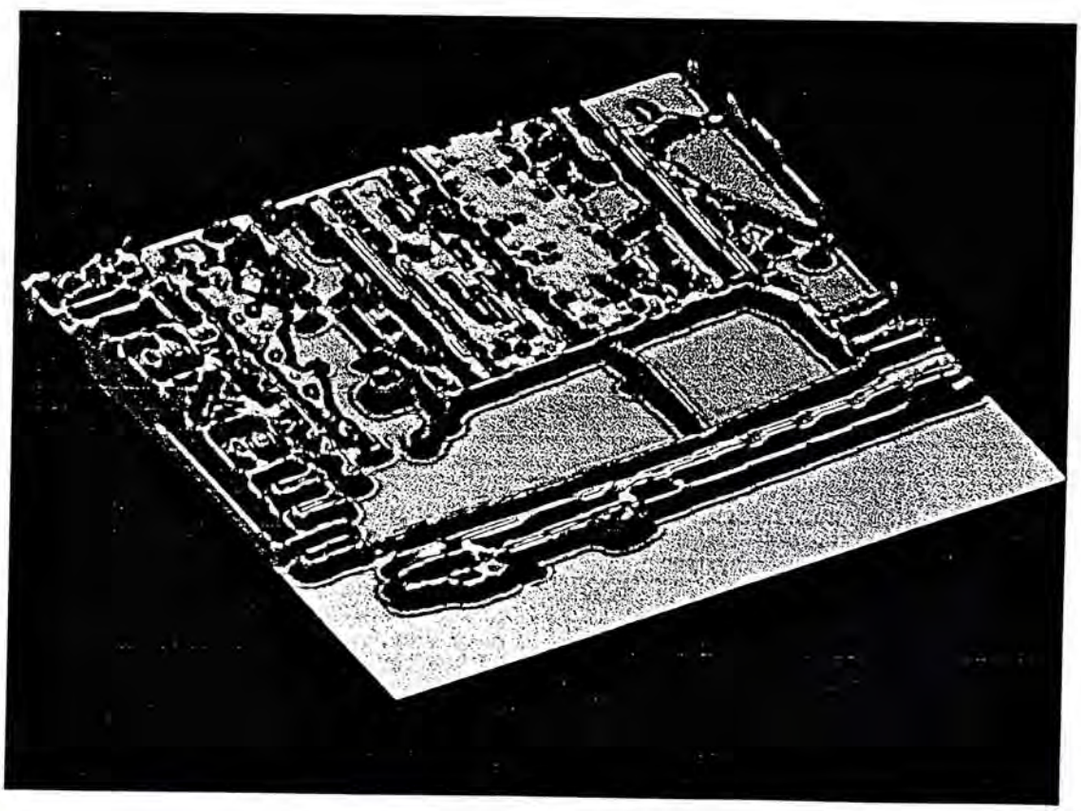
(a) The zero-crossings of the left image. (b) The zero-crossing of the right image.

Finger 4.17 The zero-crossings of the second exapmle.

As a final example, a view of the Pentagon in Washington, DC is used. The stereo pair of the images are shown in Figure 4.19. The size of these images is 512 by 512 pixels and 255 gray-level. These pair of stereo images have been used by a variety of researchers. They are a very good example to show the difficulties in the stereo vision. From Figure 4.19, it can be seen that there are many small structural details on the roof and they are too small that the edge operators can hardly to pick them up. However, they are one of the perceptual cues to human observers. The occlusions in the scene is another problem for the stereo vision researcher. The central interior of the Pentagon is at the ground level for human perception. However, it is difficult to select the right disparity for the bottom of the interior wall when it can not be seen in the another image.



(a) The disparity map of the mouse.



(b) The interpolation of the mouse.

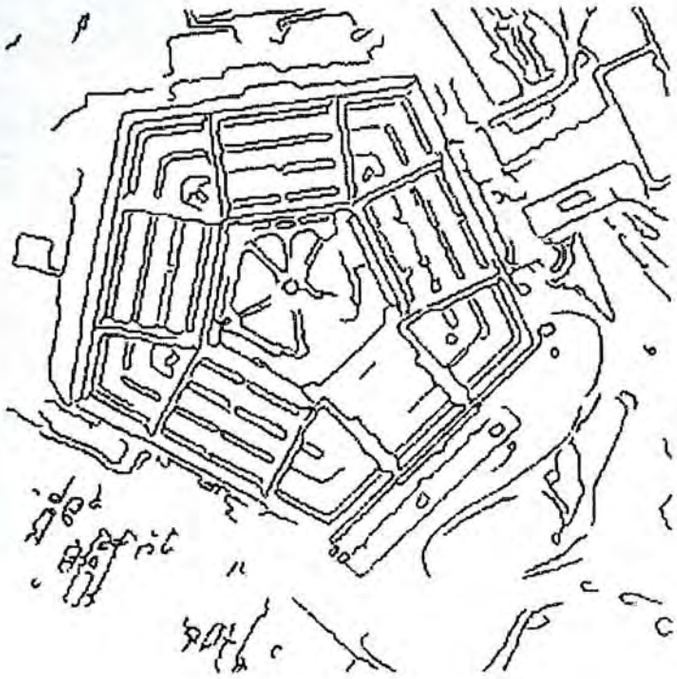
Figure 4.18 The disparity map and the disparity interpolation of the second example.

(a) The left view of the Pentagon.

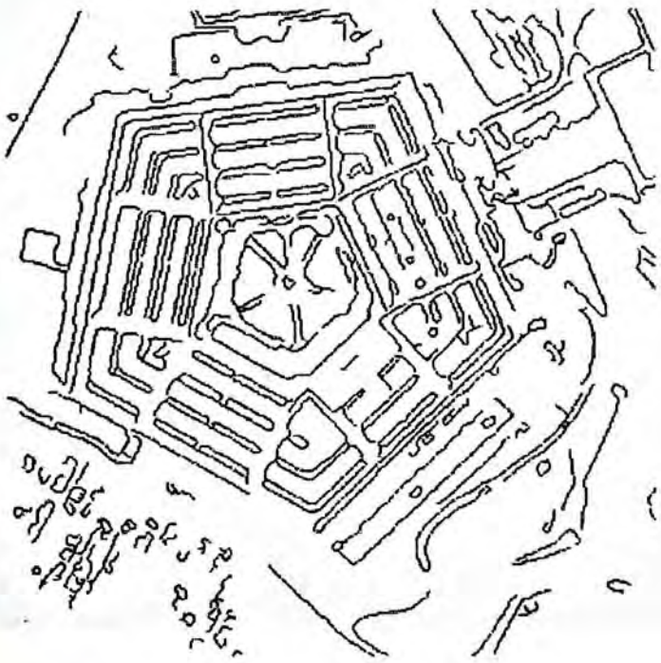
(b) The right view of the Pentagon.

Figure 4.18 Stereo images of the Pentagon.

The Canny's edge operator is used to extract the edges of these pair of images and the results are shown in Figure 4.20.

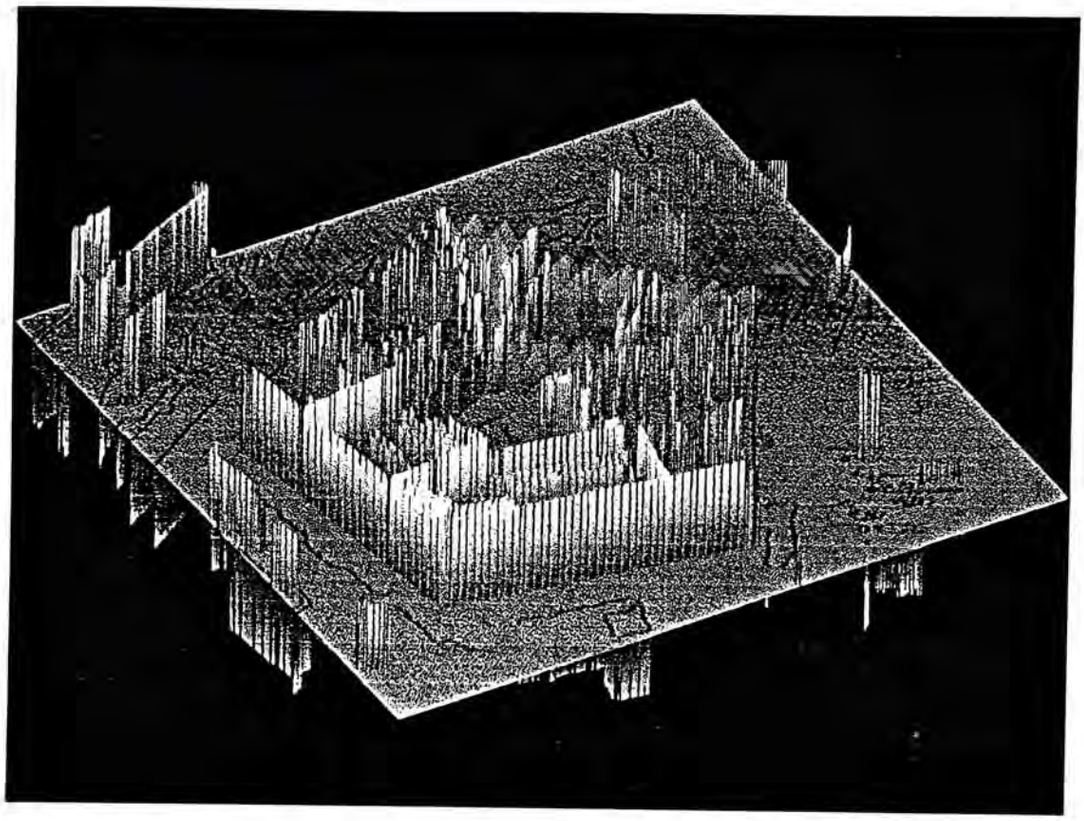


(a) The edges of the left view.

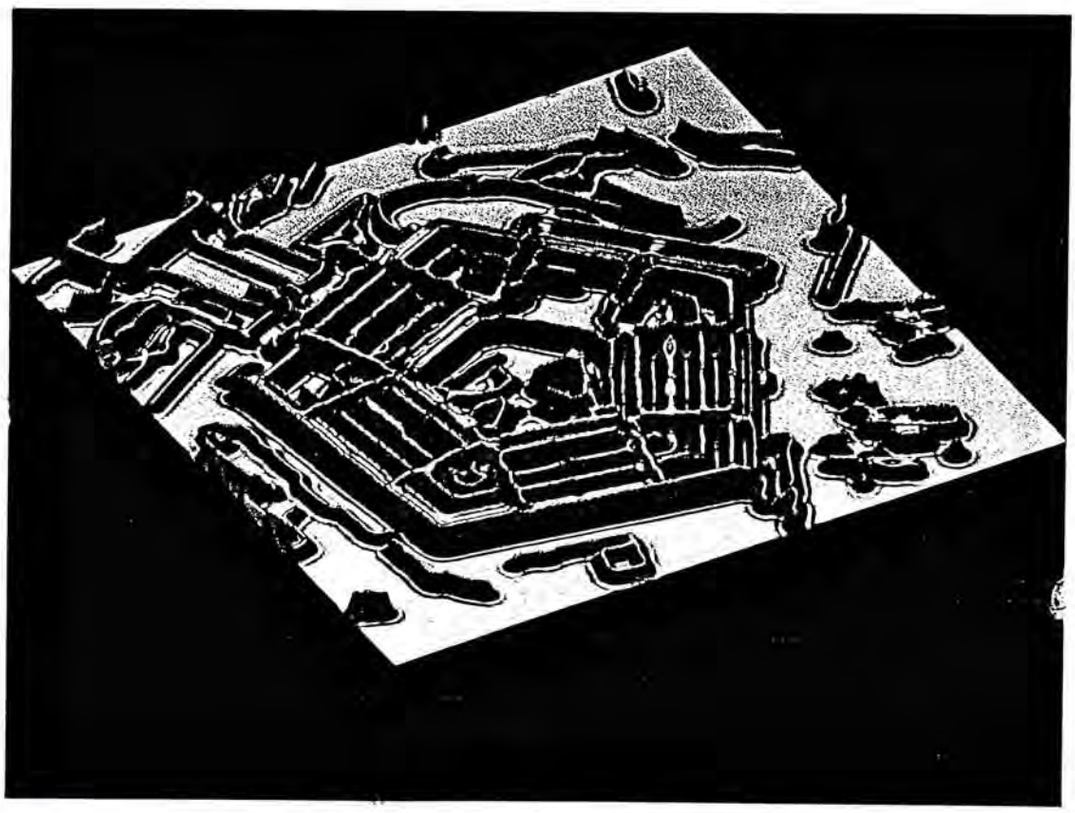


(b) The edges of the right view.

Figure 4.20 The edge detection of the Pentagon.



(a) The disparity map of the Pentagon.



(b) The interpolation of the Pentagon.

Figure 4.21 The disaprity map and the interpolation of the Pentagon.

The final disparity map obtained by PSVM is illustrated in Figure 4.21(b). The interpolation of this scene is shown in Figure 4.21(b). This disparity map is in agreement with the objects in the scene. Here, the general relief of the scene is captured by the model and the Pentagon is observed to be higher than the background. Furthermore, the center of the building is in the lower.

4.8 Discussion

In this chapter, the matching mechanism of PSVM is discussed and its performance on a series of images, including artificial image and natural images, are presented and found to be acceptable. However, in natural images, an exact evaluation of the performance of the model is difficult. It is well known that the quality of input images is one of the mean factors that affects the performance of a computer vision model. Therefore, a good input device is important to a computer vision system.

Besides the quality of images, the performance of edge detectors are important to stereo vision system. If the input image is well enough and the edge detector can locate the edges precisely, the stereo vision system will be much easy to be implemented.

4.9 Overall conclusion

The depth of images is easily measured by our human. Surprisingly, it is very difficult to develop a successful automatic stereo vision system. In this project, we extent the computer stereo vision by using psychophysical findings about human binocular and proposed a visual model for stereo vision (described in section 3.1). This model suggests that a special data structure called hypercolumn can be used as a database in binocular processing. The hypercolumn contains several columns, including location columns, orientation columns, ocular dominance columns and disparity columns. Data in the hypercolumn is a topographical mapping of the receptor surface within a three-dimensional structure. With this structure, the stereo matching can be done in parallel and the representation of a three-dimensional object is easy.

Furthermore, the visual model for stereo vision also suggests a special feature detector, simple cell, for feature extraction. The image will be passed to the feature detector after edge detection. With this scheme, it is possible to use any kinds of edge detectors.

Based on the above visual model, a computerized visual model for PSVM is developed. PSVM consists of three main stages. The first stage is local oriented line extraction. PSVM uses the feature detectors to extract on-type lines and off-type lines, and four kinds of oriented lines: horizontal lines, vertical lines, left diagonal lines and right diagonal lines. The above short lines will be placed into hypercolumn and ready for binocular matching. The second stage is local lines matching. PSVM uses the short lines in the hypercolumn to perform two kinds of matchings, short oriented lines matching and long oriented lines matching. That two matchings will give more disparity information about the scene and can be done in parallel. Note that the long oriented lines are constructed by the short oriented lines in the hypercolumn. Matching in PSVM is not restricted to one dimension. It uses fusional method to match possible primitives over an area. The third stage is for the disparity integrations. Local disparities that come from the second stage will be integrated and the final disparities are placed into hypercolumn. PSVM uses the competition model, called voter and redistributor, to select unambiguous disparity within a voting area. Once a winner is voted, it will be reassigned to this voting area. Note that the figural continuity constraint naturally implied, for PSVM uses the oriented lines as matching elements.

PSVM is tested on a number of synthetic images as well as natural images and produces satisfactory results. However, it should be pointed out that PSVM is very interested in the objects with long straight lines. It is because PSVM uses only one of the properties of simple cell, the straight line extraction function. To improve this feature extraction, more properties of the simple cell should be considered, for example, special points extraction, and the properties of the other cortical cells may also be used, for example hypercomplex cortical cells. Finally, there are other research issues such as how to use the hypercolumn to achieve interpolation. It is an interesting problem.

Appendix: Illustration example

A1 Hypercolumn structure in PSVM

In PSVM, each hypercolumn cell contains information of a straight line including, starting point coordinates, ending point coordinates, type of the line and disparity. The definition of this structure in PSVM is as follows:

```
structure hypercolumn_cell { int   endingX;
                             int   endingY;
                             char type; /* "N" for on-type line, "F" for off-type
line */
                              int  disparity;
                            }
```

A three-dimensional array is used to represent one orientation hypercolumn. Therefore, there are four kinds of this array (horizontal, vertical, left diagonal and right diagonal oriented line). The definition is

structure hypercolumn_cell hypercolumn[D][X][Y]

where D means left eye or right eye, X and Y are the coordinates of the images.

A2 Simple example

The following is a simple example to illustrate how PSVM works. The results will be shown in each stage. The input images are shown in Figure A1. The scene is composed of two square plane.

A2.1 Stage 1: Local oriented lines extraction

In this stage, PSVM extracts two type of lines, on-type lines and off-type lines, and four kinds of oriented lines, horizontal, vertical, left diagonal, and right diagonal oriented lines. The on-type lines and off-type lines are shown in Figure A2. The length of the oriented line is less than 3 pixels. Note that different oriented lines will be placed into different hypercolumn. The results of the short oriented lines are listed in Table A1. Note that no diagonal lines are detected in this example.

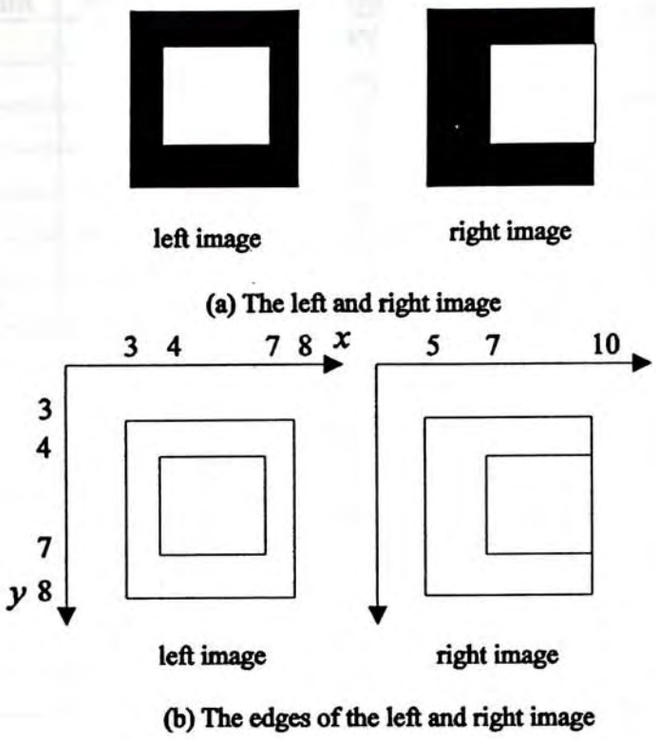


Figure A1. Input images and their edges.

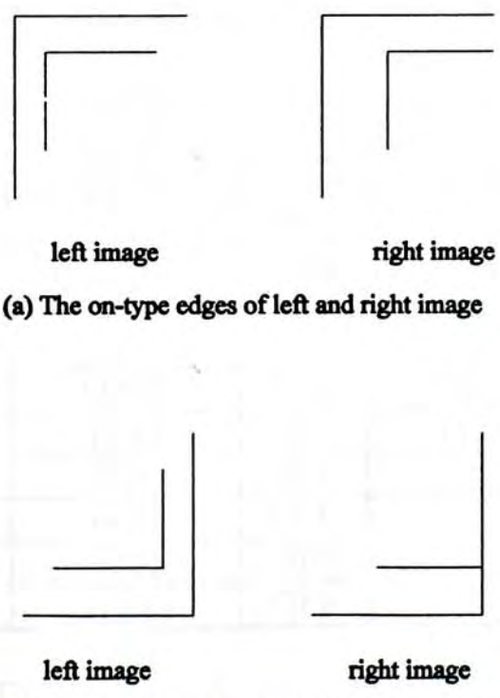


Figure A2. On-type edges and off-type edges of the input images.

	Horizontal Orientation Hypercolumn			Vertical Orientation Hypercolumn		
No.	Starting Point	Ending Point	Type	Starting Point	Ending Point	Type
SL1	(3,3)	(4,3)	ON	(3,3)	(3,4)	ON
SL2	(4,3)	(5,3)	ON	(3,4)	(3,5)	ON
SL3	(5,3)	(6,3)	ON	(3,5)	(3,6)	ON
SL4	(6,3)	(7,3)	ON	(3,6)	(3,7)	ON
SL5	(7,3)	(8,3)	ON	(3,7)	(3,8)	ON
SL6	(4,4)	(5,4)	ON	(4,4)	(4,5)	ON
SL7	(5,4)	(6,4)	ON	(4,5)	(4,6)	ON
SL8	(6,4)	(7,4)	ON	(4,6)	(4,7)	ON
SL9	(4,7)	(5,7)	OFF	(7,4)	(7,5)	OFF
SL10	(5,7)	(6,7)	OFF	(7,5)	(7,6)	OFF
SL11	(6,7)	(7,7)	OFF	(7,6)	(7,7)	OFF
SL12	(3,8)	(4,8)	OFF	(8,3)	(8,4)	OFF
SL13	(4,8)	(5,8)	OFF	(8,4)	(8,5)	OFF
SL14	(5,8)	(6,8)	OFF	(8,5)	(8,6)	OFF
SL15	(6,8)	(7,8)	OFF	(8,6)	(8,7)	OFF
SL16	(7,8)	(8,8)	OFF	(8,7)	(8,8)	OFF

(a) The short oriented lines of the left image.

	Horizontal Orientation Hypercolumn			Vertical Orientation Hypercolumn		
No.	Starting Point	Ending Point	Type	Starting Point	Ending Point	Type
SR1	(5,3)	(6,3)	ON	(5,3)	(5,4)	ON
SR2	(6,3)	(7,3)	ON	(5,4)	(5,5)	ON
SR3	(7,3)	(8,3)	ON	(5,5)	(5,6)	ON
SR4	(8,3)	(9,3)	ON	(5,6)	(5,7)	ON
SR5	(9,3)	(10,3)	ON	(5,7)	(5,8)	ON
SR6	(7,4)	(8,4)	ON	(7,4)	(7,5)	ON
SR7	(8,4)	(9,4)	ON	(7,5)	(7,6)	ON
SR8	(9,4)	(10,4)	ON	(7,6)	(7,7)	ON
SR9	(7,7)	(8,7)	OFF	(10,3)	(10,4)	OFF
SR10	(8,7)	(9,7)	OFF	(10,4)	(10,5)	OFF
SR11	(9,7)	(10,7)	OFF	(10,5)	(10,6)	OFF
SR12	(5,8)	(6,8)	OFF	(10,6)	(10,7)	OFF
SR13	(6,8)	(7,8)	OFF	(10,7)	(10,8)	OFF
SR14	(7,8)	(8,8)	OFF			
SR15	(8,8)	(9,8)	OFF			
SR16	(9,8)	(10,8)	OFF			

(b) The short oriented lines of the right image

Table A1. The results of the stage 1 (Local oriented lines extraction).

A2.2 Stage 2: Local lines matching

There are two matching processes, short oriented lines matching and long oriented lines matching, in Stage 2. The short oriented lines matching directly uses the information in hypercolumn for matching while the long oriented lines matching will reconstruct the short oriented lines in hypercolumn and uses them for matching. The short oriented lines matching results are listed in Table A2. The results of long oriented lines after reconstruction are shown in Table A3 and their matching results are listed in Table A4.. Note that the matching results are placed into the left hypercolumn. It should point out that SL9 does not match SR6 in short oriented lines matching for the type is not same (see Table A1). And in long oriented lines matching, LL3 does not match LR3 for the fusional value (*fp*) is large then that of LL4→LR3 (see section 4.4.1).

A2.3 Stage 3: Disparity integration

Disparities found in Stage 2 are integrated at Stage 3. In this stage, PSVM selects the most popular disparity within a voting area. Once the popular disparity is selected, it will be assigned to that area. The results of this stage are listed in Table A5.

	Horizontal Orientation Hypercolumn		Vertical Orientation Hypercolumn	
No.	Disparity	Remark	Disparity	Remark
SL1	-2	SL1→SR1	-2	SL1→SR1
SL2	-2	SL2→SR2	-2	SL2→SR2
SL3	-2	SL3→SR3	-2	SL3→SR3
SL4	-2	SL4→SR4	-2	SL4→SR4
SL5	-2	SL5→SR5	-2	SL5→SR5
SL6	-3	SL6→SR6	-3	SL6→SR6
SL7	-3	SL7→SR7	-3	SL7→SR7
SL8	-3	SL8→SR8	-3	SL8→SR8
SL9	-3	SL9→SR9	-3	SL9→SR10
SL10	-3	SL10→SR10	-3	SL10→SR11
SL11	-3	SL11→SR11	-3	SL11→SR12
SL12	-2	SL12→SR12	-2	SL12→SR9
SL13	-2	SL13→SR13	Null	No Match
SL14	-2	SL14→SR14	Null	No Match
SL15	-2	SL15→SR15	Null	No Match
SL16	-2	SL16→SR16	-2	SL16→SR13

Table A2. The results of the short oriented lines matching.

	Horizontal Orientation Hypercolumn			Vertical Orientation Hypercolumn		
No.	Starting Point	Ending Point	Type	Starting Point	Ending Point	Type
LL1	(3,3)	(8,3)	ON	(3,3)	(3,8)	ON
LL2	(4,4)	(7,4)	ON	(4,4)	(4,7)	ON
LL3	(4,7)	(7,7)	OFF	(7,4)	(7,7)	OFF
LL4	(3,8)	(8,8)	OFF	(8,3)	(8,8)	OFF

(a) The long oriented lines of the left image after reconstruction.

	Horizontal Orientation Hypercolumn			Vertical Orientation Hypercolumn		
No.	Starting Point	Ending Point	Type	Starting Point	Ending Point	Type
LR1	(5,3)	(10,3)	ON	(5,3)	(5,8)	ON
LR2	(7,4)	(10,4)	ON	(7,4)	(7,7)	ON
LR3	(7,7)	(10,7)	OFF	(10,3)	(10,8)	OFF
LR4	(5,8)	(10,8)	OFF			

(b) The long oriented lines of the right image after reconstruction.

Table A3. The results of long oriented lines reconstruction.

	Horizontal Orientation Hypercolumn		Vertical Orientation Hypercolumn	
No.	Disparity	Remark	Disparity	Remark
LL1	-2	LL1→LR1	-2	LL1→LR1
LL2	-3	LL2→LR2	-3	LL2→LR2
LL3	-3	LL3→LR3	Null	No Match
LL4	-2	LL4→LR4	-2	LL4→LR3

Table A4. The matching results of the long oriented lines.

	Horizontal Orientation Hypercolumn				Vertical Orientation Hypercolumn			
No.	Starting Point	Ending Point	Type	Disparity	Starting Point	Ending Point	Type	Disparity
SL1	(3,3)	(4,3)	ON	-2	(3,3)	(3,4)	ON	-2
SL2	(4,3)	(5,3)	ON	-2	(3,4)	(3,5)	ON	-2
SL3	(5,3)	(6,3)	ON	-2	(3,5)	(3,6)	ON	-2
SL4	(6,3)	(7,3)	ON	-2	(3,6)	(3,7)	ON	-2
SL5	(7,3)	(8,3)	ON	-2	(3,7)	(3,8)	ON	-2
SL6	(4,4)	(5,4)	ON	-3	(4,4)	(4,5)	ON	-3
SL7	(5,4)	(6,4)	ON	-3	(4,5)	(4,6)	ON	-3
SL8	(6,4)	(7,4)	ON	-3	(4,6)	(4,7)	ON	-3
SL9	(4,7)	(5,7)	OFF	-3	(7,4)	(7,5)	OFF	-3
SL10	(5,7)	(6,7)	OFF	-3	(7,5)	(7,6)	OFF	-3
SL11	(6,7)	(7,7)	OFF	-3	(7,6)	(7,7)	OFF	-3
SL12	(3,8)	(4,8)	OFF	-2	(8,3)	(8,4)	OFF	-2
SL13	(4,8)	(5,8)	OFF	-2	(8,4)	(8,5)	OFF	-2
SL14	(5,8)	(6,8)	OFF	-2	(8,5)	(8,6)	OFF	-2
SL15	(6,8)	(7,8)	OFF	-2	(8,6)	(8,7)	OFF	-2
SL16	(7,8)	(8,8)	OFF	-2	(8,7)	(8,8)	OFF	-2

Table A5. The final disparity of the simple example.

References

[1] Baker, H. H. "Edge based stereo correlation," *Proceedings of Image Understanding Workshop*, Colledge Park, Md., Apr., 1980, pp. 168-175.

[2] Baker, H. H and Binford, T. O. "Depth from edge and intensity stereo," *Proc., 7th Int. Joint Conf. Artificial Intelligence*, 1981, pp. 631-636.

[3] Barlow, H., Blakemore, C., Pettigrew, J. D. "The neural mechanism of binocular depth discrimination," *J. Physiol. (Lond.)*, 193, 1967, pp. 327-342.

[4] Canny, J. F. "Finding edges and lines in images," Massachusetts Institute of Technology Artificial Intelligence Laboratory Technical Report Tr-720, June 1983.

[5] Charniak, E. and McDermott, D. *Introduction to Artificial Intelligence*, Addison-Wesley, 1985.

[6] Clarke, P. G. H. and Whitteridge, D. "The cortical visual area of the sheep," *J. Physiol. (Lond.)*, 256, 1976, pp. 497-508.

[7] Cox, Ingemar J., Hingorani, Sunita , Maggs, Bruce M., and Rao, Satish B. "Stereo without disparity gradient-smoothing: a Bsyesian sensor fusion solution," *Brithsh Machine Vision Conf., Leeds*, England, Sep., 1992, pp. 337-346.

[8] Ferster, D. "A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex," *J. Physiol. (Lond.)*, 311, 1980, pp. 623-655.

[9] Goldstein, E. B. *Sensation and perception*, third edition, Wadsworth 1989.

[10] Grimson. W. E. L., *From Image to Surfaces,* MIT Press, 1981.

[11] Grimson. W. E. L. "Computing stereopsis using feature point contour matching," *Techniques for 3-D Machine Perception,* Elsevier Science, 1986, pp. 75-111.

[12] Griswold, N. C. and Teh, C. P. "A new stereo vision model based upon the binocular fusion concept," *Comput. Vision, Graphics, and Image Processing*, 41, 1988, pp. 153-171.

[13] Hirai, Y. and Fukushima, K. "A model of neural network extracting binocular parallax," *Biol. Cybernetics* 18, 1975, pp. 19-29.

[14] Howard, S. *Binocular vision-a programmed text,* Heinemann Press, 1981.

[15] Hsieh, Y. C., McKeown, D. M., and Perlant, P. "Performance evaluation of scene registration and stereo matching for cartographic feature extraction," *IEEE, Trans. Patt. Anal. Machine Intell.*, vol. 14, No. 2, Feb. 1992, pp. 214-237.

[16] Hubel, D. H. and Wiesel, T. N. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol. (Lond.)*, 160, 1962, pp. 106-154.

[17] Hubel, D. H. and Wiesel, T. N. "Cells sensitive to binocular depth in area 18 of the macaque monkey cortex," *Nature*, 225, 1970, pp. 41-42.

[18] Hubel, D. H. and Wiesel, T. N. "Functional architecture of macaque monkey visual cortex," *Proc. R. Soc. Lond.*, B 196, 1977, pp. 1-59.

[19] Kaplan, E. "The receptive field structure of retinal ganglion cells in cat and monkey," *Vision and Visual Dysfunction: Vol. 4, The Neural Basis of Visual Function*, Macmillan Press, 1991, pp. 10-40.

[20] Kass, M. "A computational framework for the visual correspondence problem," *Proc. 8th Int. Joint Conf. Artificial Intell.*, Aug., 1983, pp. 1043-1045.

[21] LeVay, S., and Nelson, S. B. "Columnar organization of the visual

cortex," *Vision and Visual Dysfunction: Vol. 4, The Neural Basis of Visual Function*, Macmillan Press, 1991, pp. 266-315.

[22] Marr, D. and Poggio, T. "A computation theory of human stereo vision," *Proc. R. Soc. Lond.*, B 204, 1979, pp. 301-328.

[23] Marr, D., and Hildreth, E. "Theory of edge detection," *Proc. R. Soc. Lond.*, B 207, 1980, pp. 187-217.

[24] Marr, D. *Vision*, Freeman, San Francisco, 1982.

[25] Mayhew, J. E. V. and Frisby, J. P. "Psychophysical and computational studies towards a theory of human stereopsis," *Artificial Intelligence*, 17, 1981, pp. 349-385.

[26] Medioni, G., and Nevatia, R. "Segment-based stereo matching, Comput. Vision," *Graphics, Image Processing*, vol. 31, July, 1985, pp. 2-18.

[27] Ohta, Y. and Kanade, T. "Stereo by intra- and inter-scankine search using dynamic programming," *IEEE Trans., Pattern Analysis and Machine Intelligence*, Vol. PAM-7, No. 2, March 1985, pp. 301-328.

[28] Olsen, S. L. "Stereo correspondence by surface reconstruction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 12, No. 3, 1990, pp. 309-315.

[29] Or, S. H. and Wong, K. H., "A multi-layer global competition approach for stereo disparity computation," *Proceedings of the 2nd Pacific Rim International Conference on Artificial Intelligence*, 1992, pp. 888-895.

[30] Poggio, G. F. and Poggio, T. "The analysis of stereopsis," *Ann. Rev. Neurosci.*, 7, 1984, pp. 379-412.

[31] Richard, P. W. "Computational vision with reference to binocular stereo vision," *Science of Vision*, Springer-Verlag, 1990, pp. 332-364.

[32] Rodieck, E. "Analysis of receptive fields of cat retinal ganglion cell," *J. Neurophysiol.* 28, 1965, pp. 833-849.

[33] Shaw, T. "Local and regional edge detectors: some comparisons,"

Computer Graphics and Image Processing, 9, 1979, pp. 135-149.

[34] Tyler, C. W. and Scott, A. B. "Binocular vision," Physiology of the Human eye and Visual System, Harper & Row, 1979, pp. 643-671.

[35] Tyler, C. W. "The horopter and binocular fusion," *Vision and Visual Dysfunction: Vol. 9, Binocular Vision*, Macmillan Press, 1991, pp. 19-37.